# DOCUMENT RESUME

| | |
|---|---|
| ED 111 748 | SO 008 606 |

| | |
|---|---|
| AUTHOR | Fair, Jean, Ed. |
| TITLE | National Assessment and Social Studies Education: A Review of Assessments in Citizenship and Social Studies by the National Council for the Social Studies. |
| INSTITUTION | Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.; National Council for the Social Studies, Washington, D.C. |
| SPONS AGENCY | Carnegie Corp. of New York, N.Y.; Ford Foundation, New York, N.Y.; Office of the Assistant Secretary for Education (DHEW), Washington, D.C. |
| PUB DATE | 75 |
| NOTE | 121p.; For the full NCSS report, See ED 100 739-746 |
| AVAILABLE FROM | Superintendent of Documents, U. S. Government Printing Office, Washington, D.C. 20402 ($1.90) |
| EDRS PRICE | MF-$0.76 HC-$5.70 Plus Postage |
| DESCRIPTORS | *Citizenship; *Educational Assessment; *Educational Objectives; *Evaluation; Evaluation Methods; Measurement Goals; Measurement Instruments; Models; National Surveys; *Social Studies; Test Validity |
| IDENTIFIERS | *National Assessment |

ABSTRACT

This report of a study conducted by the National Council for the Social Studies (NCSS) reviews, interprets, and disseminates findings of the National Assessment of Educational Progress (NAEP) assessments in citizenship and social studies. The study represents an investigation funded by the Education Commission of the States and coordinated by a special steering committee working under the auspices of NCSS. Chapter titles include: (1) National Assessment and Social Studies Education: The Setting, by Jean Fair; (2) Potential Uses of the National Assessment Model at the State Level and for Accountability Purposes, by Bob L. Taylor; (3) On the Need for Criterion-Referenced Research and Demonstration: A Reaction to the Model of National Assessment in Citizenship, by Joseph C. Grannis; (4) Evaluating Social Studies and Citizenship Education: Some Alternate Approaches, by Michael Scriven; (5) Critique of NAEP Objectives and Procedures: Citizenship and Social Studies, by A. Guy Larkins; (6) Validity of Social Studies and Citizenship Exercises, by Francis P. Hunkins; (7) A Rating of Social Studies Exercises by Social Studies Educators, by June Chapin; (8) An Analysis of a Selected Set of Social Studies Exercises: Knowledge of Institutions, by C. Benjamin Cox; and (9) A Response from the National Assessment, by J. Stanley Ahmann. The appendixes include the NAEP citizenship and social studies objectives, performance evaluations, and a bibliography. (DE)

ED111748

# NATIONAL ASSESSMENT
# AND
# SOCIAL STUDIES EDUCATION

## A Review of Assessments in Citizenship and Social Studies by the National Council for the Social Studies

Jean Fair, Editor

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SO 008 606

For Sale by the Superintendent of Documents, U.S. Government Printing Office, Washington D.C. 20402

$1.90

National Assessment of Educational Progress, 1975

*This publication was prepared and produced pursuant to agreements with the National Center for Education Statistics of the Office of the Assistant Secretary for Education, Department of Health, Education, and Welfare, with additional funds from the Carnegie Corporation of New York and the Ford Foundation's Fund for the Advancement of Education. The statements and views expressed herein do not necessarily reflect the position and policy of the U.S. Office of Education or other grantors but are solely the responsibility of the National Assessment of Educational Progress, a project of the Education Commission of the States.*

A Report from the

**National Council for the Social Studies**
**1201 Sixteenth Street, N.W.**
**Washington, D.C. 20036**

prepared with the assistance of the

**National Assessment of Educational Progress**
**(A Project of the Education Commission of the States)**
**1860 Lincoln Street, Suite 700**
**Denver, Colorado 80203**

---

*The full NCSS report (S0008019-026) is available through the Educational Resources Information Center (ERIC) Document Reproduction Service, Box 190, Arlington, VA 22210. It can also be found in microfiche in state education department libraries and more than 500 college and university libraries.*

# TABLE OF CONTENTS

LIST OF EXHIBITS . . . . . . . . . . . . . . . . . . . . . . . . v

LIST OF TABLES . . . . . . . . . . . . . . . . . . . . . . . . vii

FOREWORD . . . . . . . . . . . . . . . . . . . . . . . . ix

NCSS STUDY PARTICIPANTS . . . . . . . . . . . . . . . . . . . . . . . . xi

CHAPTER 1. National Assessment and Social Studies Education: The Setting *by Jean Fair* . . . . . . . . . . . . . . . . . . 1

CHAPTER 2. Potential Uses of the National Assessment Model at the State Level and for Accountability Purposes *by Bob L. Taylor* . . . . . . . . 5
- The Model . . . . . . . . . . . . . . . . . . . . . . . . 5
- Use of the Model at State Level . . . . . . . . . . . . . . . . . 9
- Implications of the Model for Curriculum Development . . . . . . . . . . 11
- Use of the Model for Accountability Purposes . . . . . . . . . . . . . 13
- Summary . . . . . . . . . . . . . . . . . . . . . . . . 14

CHAPTER 3. On the Need for Criterion-Referenced Research and Demonstration: A Reaction to the Model of National Assessment in Citizenship *by Joseph C. Grannis* . . . . . . . . . . . . . . . . . . 17
- The National Assessment Model as Representation and as Influence . . . 17
- The Meaning of Patterns in National Assessment Findings . . . . . . . . 19
- An Alternative to the National Assessment Model . . . . . . . . . . . . 23

CHAPTER 4. Evaluating Social Studies and Citizenship Education: Some Alternate Approaches *by Michael Scriven* . . . . . . . . . . 27
- Alternatives to National Assessment . . . . . . . . . . . . . . . . 27
- Changes in National Assessment . . . . . . . . . . . . . . . . . 29
- Comments on the Taylor Chapter . . . . . . . . . . . . . . . . . 30
- Conclusion . . . . . . . . . . . . . . . . . . . . . . . . 31

CHAPTER 5. Critique of NAEP Objectives and Procedures: Citizenship and Social Studies *by A. Guy Larkins* . . . . . . . . . . . . . . . 33
- Objectives . . . . . . . . . . . . . . . . . . . . . . . . 33
- Exercise Development . . . . . . . . . . . . . . . . . . . . 35
- Sampling and Data Analysis . . . . . . . . . . . . . . . . . . 36
- Conclusion . . . . . . . . . . . . . . . . . . . . . . . . 37

CHAPTER 6. Validity of Social Studies and Citizenship Exercises *by Francis P. Hunkins* . . . . . . . . . . . . . . . . . . . . . . . . 39
- Social Studies, Citizenship—Goals and Objectives of National Assessment . . . . . . . . . . . . . . . . . . . . . . . . 39
- Specific Nature of the Task . . . . . . . . . . . . . . . . . . 40
- The Results . . . . . . . . . . . . . . . . . . . . . . . . 40
- Final Comments . . . . . . . . . . . . . . . . . . . . . . . . 44

CHAPTER 7. A Rating of Social Studies Exercises by Social Studies Educators *by June Chapin* . . . . . . . . 45
- Compatibility with NCSS Curriculum Guidelines . . . . . . 45
- Desirability of the Exercises . . . . . . 46
- Realistically Satisfactory Performance . . . . . . 46
- Summary . . . . . . 47

CHAPTER 8. An Analysis of a Selected Set of Social Studies Exercises: Knowledge of Institutions *by C. Benjamin Cox* . . . . . . 49
- Introduction . . . . . . 49
- A Review of the Data in the National Assessment of Social Studies, 1971-72 . . . . . . 54
- Comparison of Group Differences from National Success Levels in Social Studies and Citizenship Assessments . . . . . . 58
- Analysis of the Social Studies Data by Knowledge of Structures and by Knowledge of Rights and Duties . . . . . . 64
- Summary . . . . . . 92

CHAPTER 9. A Response from National Assessment *by J. Stanley Ahmann* . . . 95

APPENDIX A. Citizenship and Social Studies Objectives (First Assessment) . . 99

APPENDIX B. Revised Citizenship and Social Studies Objectives . . . . . . 101

APPENDIX C. Evaluation of Performance on Selected Social Studies Exercises 105

APPENDIX D. Bibliography of National Assessment Publications Cited . . . . 115

# LIST OF EXHIBITS

EXHIBIT 1. Components of the Model . . . . . . 6

EXHIBIT 2. Categorical Breakdown of NAEP Assessment Data . . . . . . 53

EXHIBIT 3. Knowledge of Institutions—Median Raw P-Values—Nation, Region, Race, STOC, Parental Education—Social Studies 1971-72 . . . . . . 55

EXHIBIT 4. Knowledge of Institutional Structures—Median Raw P-Values—Nation, Region, Race, STOC, Parental Education—Social Studies 1971-72 . . . . . . 56

EXHIBIT 5. Knowledge of Institutional Rights and Duties—Median Raw P-Values—Nation, Region, Race, STOC, Parental Education—Social Studies 1971-72 . . . . . . 57

EXHIBIT 6. Knowledge of Institutions—Regions—Median Delta P . . . . . . 59

EXHIBIT 7. Knowledge of Institutions—Regions—Social Studies 1971-72 . . 60

EXHIBIT 8. Knowledge of Governmental Structures and Functions—Regions—Citizenship 1969-70 . . . . . . 61

EXHIBIT 9. Knowledge of Institutions—Race—Social Studies 1971-72 . . . . 62

EXHIBIT 10. Knowledge of Governmental Structures and Functions—Race—Citizenship 1969-70 . . . . . . 63

EXHIBIT 11. Knowledge of Institutions—Parental Education—Social Studies 1971-72 . . . . . . 65

EXHIBIT 12. Knowledge of Governmental Structures and Functions—Parental Education—Citizenship 1969-70 . . . . . . 66

EXHIBIT 13. Knowledge of Institutions—STOC—Social Studies 1971-72 . . . 67

EXHIBIT 14. Knowledge of Governmental Structures and Functions—STOC—Citizenship 1969-70 . . . . . . 68

EXHIBIT 15. Knowledge of Institutional Structures—Regions—Social Studies 1971-72 . . . . . . 70

EXHIBIT 16. Knowledge of Institutional Rights and Duties—Regions—Social Studies 1971-72 . . . . . . 71

EXHIBIT 17. Knowledge of Institutional Structures—Sex—Social Studies 1971-72 . . . . . . 74

EXHIBIT 18. Knowledge of Institutional Rights and Duties—Sex—Social Studies 1971-72 . . . . . . 76

EXHIBIT 19. Knowledge of Institutional Structures—Race—Social Studies 1971-72 . . . . . 78
EXHIBIT 20. Knowledge of Institutional Rights and Duties—Race—Social Studies 1971-72 . . . . . 80
EXHIBIT 21. Knowledge of Institutional Structures—Parental Education—Social Studies 1971-72 . . . . . 83
EXHIBIT 22. Knowledge of Institutional Rights and Duties—Parental Education—Social Studies 1971-72 . . . . . 85
EXHIBIT 23. Knowledge of Institutional Structures—STOC—Social Studies 1971-72 . . . . . 88
EXHIBIT 24. Knowledge of Institutional Rights and Duties—STOC—Social Studies 1971-72 . . . . . 90

# LIST OF TABLES

TABLE 1. Summary of Social Studies Released and Unreleased Exercises . . 41

TABLE 2. Summary of Citizenship Released and Unreleased Exercises . . . 42

TABLE 3. Summary of NAEP Social Studies Exercises Classified by General Desirability . . . . . . . . . . . . . . . . . . . . . . . . . 46

TABLE 4. Summary of NAEP Social Studies Exercises Classified by National Level of Realistically Satisfactory Performance . . . . 47

TABLE 5. Sample Exercise Data—13-Year-Old Respondents . . . . . . . . . 50

TABLE 6. Color-Scheme Display of Exercises and Performances . . . . . . . 51

TABLE 7. Exercise Distribution in Rights and Duties Color Scheme . . . . . 52

TABLE 8. Exercise Distribution in Knowledge of Structures Color Scheme . 52

TABLE 9. Northeast/Southeast/National Comparisons on Selected Exercises 72

TABLE 10. Comparison of High and Low Metro and Extreme Rural Adults on Selected Rights and Duties Exercises—Social Studies 1971-72, Delta P . . . . . . . . . . . . . . . . . . . 92

# FOREWORD

National Assessment of Educational Progress (NAEP), a project of the Education Commission of the States, is a research effort designed to gather information about the educational achievements of 9-year-olds, 13-year-olds, 17-year-olds and adults (ages 26-35) in 10 learning areas. Different learning areas are assessed every year and all areas are periodically reassessed in order to measure educational change. Plans for assessment efforts in the next five years are outlined in the final chapter of this report.

Because of its focus on the collection and analysis of data, National Assessment must rely on professional educators and organizations to study that data and draw out implications and meaning. The project welcomes the serious interpretation of its findings and realizes that only through such follow-up work and dissemination can the results have a bearing on education decisions and plans. We therefore welcome this report — the result of a grant from NAEP to the National Council for the Social Studies (NCSS). The council organized for its task (see Chapter 1) and called on a number of its members to participate. National Assessment and NCSS are grateful to the many professional educators, listed on the following pages, who contributed to the study and this review.

J. STANLEY AHMANN
*Director*
National Assessment of Educational Progress

# NCSS STUDY PARTICIPANTS

## Steering Committee

Jean Fair
College of Education
Wayne State University

Dana Kurfman
Social Studies Curriculum Chairman
Prince George County, Maryland Public Schools

June Gilliard
Joint Council on Economic Education
New York City

James Shaver
Bureau of Educational Research
College of Education
Utah State University

Ronald Smith
Curriculum Consultant
Portland (Oregon) Public Schools

## Investigators (Authors)

Bob L. Taylor — Task 1
College of Education
University of Colorado

Guy Larkins — Task 2
College of Education
University of Georgia

Francis Hunkins — Task 3
College of Education
University of Washington

Benjamin Cox — Task 4
College of Education
University of Illinois

June Chapin — Task 5
College of Notre Dame
Belmont, California

## Task 1 — Writers of Commissioned Papers

Joseph Grannis
Teachers College
Columbia University

Michael Scriven
Department of Philosophy
University of California at Berkeley

## Task 2 — Review Panel

Ambrose Clegg
College of Education
Kent State University

Dell Felder
College of Education
University of Houston

Martha John
College of Education
Boston University

Kay Kamin
Department of Education
Rosary College
River Forest, Illinois

Robert Klingenfus
Social Studies Supervisor
Tucson (Arizona) Public Schools

Estelle Lit
Kennedy High School
Granada Hills, California

## Task 3 — Review Panel

Theresa Geiger
Instructional Specialist
Portland (Oregon) Public Schools

Michael Hartoonian
Social Studies Specialist
State Department of Public Instruction
Madison, Wisconsin

Jack Thompson
Social Studies Coordinator
Renton (Washington) Public Schools

Arthur Ware
Social Studies Coordinator
Bellevue (Washington) Public Schools

## Task 4 — Review Panel

Willard Bill
College of Education
University of Washington

Louise Boedeker
Department of Curriculum and Teaching
Hunter College, New York City

Myrtle Fentress
Curriculum Specialist
Prince George County (Maryland) Public Schools

William Guardia
Social Studies Supervisor
San Antonio (Texas) Public Schools

Edward Martin
Educational Development Center
Cambridge, Massachusetts

Richard Wisniewski
College of Education
University of Wisconsin-Milwaukee

## Task 5 — Review Panel

Charles Beaty
Social Studies Director
Shawnee Mission (Kansas) Public Schools

Y. Arturo Carera
School of Education
University of Colorado

Linda Falkenstein
Curriculum Consultant
Multnomah County (Oregon) Schools

Patricia Glasheen
Rhode Island College

Florence Jackson
Bureau of Social Studies
Board of Education, New York City

Zada Kobles
Jefferson Junior High School
Minneapolis, Minnesota

Ann Powers
Social Studies Consultant
Durham, North Carolina

Roosevelt Ratliff
Multi-Ethnic Studies
Tulsa (Oklahoma) Public Schools

Jan Tucker
College of Education
Florida International University

# CHAPTER 1

# NATIONAL ASSESSMENT AND SOCIAL STUDIES EDUCATION: THE SETTING

Jean Fair

Wayne State University

In the midst of unsettling social change, education becomes inevitably a matter of public debate and policy making. The heat, even outright turmoil, of recent years has subsided, and faith in education as the road to salvation has given way. Still firm is the belief that education is basic to individual and social welfare. And still widespread is the uneasy feeling that schools are not doing what they ought to be doing, and not even doing well what they have long been doing. When too little money is at hand for public services, the debate is further sharpened.

The pressures in policy making make plain the need for information. Simply stated, young people go to school to learn something. Although the country has accumulated information on scores of no-doubt useful matters, little systematic evidence has been available on the crucial point of it all, actual educational attainment.

Efforts to gather evidence soon confront the basic questions of the debate: what should young people be learning, what should they achieve and to what extent are they doing so? To these two are added others: how to find out and how to explain what is found in some way useful to decision making. Information gathering is no simple task: indeed, it is a subject of debate in itself.

A major effort has been under way by the National Assessment of Educational Progress (NAEP). Of special significance to social studies education are its recent reports of findings in citizenship and social studies, 1969-70 and 1971-72 assessments.

Some few words of background are needed here. Formal discussion of the possibilities of a national assessment of educational attainments began some 10 years ago. Supported by the Carnegie Corporation of New York, a private foundation, the Exploratory Committee on Assessing the Progress of Education came into being in 1964 to develop a concrete plan. This committee's work became the basis of the present national project. In mid-1969 the Education Commission of the States (ECS), a compact presently of some 47 states and territories to consider and coordinate efforts and problems, assumed the governance of the project.

The prime purpose of the National Assessment of Educational Progress is to make information available to those interested in education. Assessments are carried on at regular intervals, at the present in 10 areas: reading, writing, science, mathematics, social studies, citizenship, career and occupational development, literature, art and music — in short, in much more than the three R's. NAEP assesses educational achievement, not merely school achievement. Obviously schools have responsibilities for education. Still television, magazines, libraries, newspapers, civic organizations (especially those for young people), religious institutions, personal opportunities and experiences—these and others all contribute to education. National Assessment does not aim to distinguish one source of achievement from another.

Moreover, the Assessment is not a national examination, a set of hurdles that students must pass over for continuing opportunity. Neither does National Assessment attempt to measure the performance of any one person, school district or even state, nor to award praise or blame to any institution. Although no search can be made except from some frame of reference, NAEP consistently refrains from interpretive explanations of the data collected. Even though perfection is never to be expected nor debate cut off, National Assessment is a serious, highly professional undertaking. What can be learned from its efforts deserves attention.

Consequently, the National Council for the Social Studies (NCSS) welcomed the opportunity for independent study, interpretation and dissemination of the assessments in citizenship and social studies, two sides of the coin of social studies education. That NAEP supported this study with a grant of funds is a mark of its professionalism.

An NCSS Steering Committee took over all responsibility: Jean Fair, chairperson; June Gilliard; Dana Kurfman; James Shaver; and Ronald Smith. The task of dissemination has been the responsibility of the Steering Committee. Five other tasks were identified, and major investigators were appointed to give time and thought to examining them: (1) the assessment model, Bob Taylor; (2) methods and procedures, Guy Larkin; (3) the validity of the exercises, Francis Hunkins; (4) interpretation of the findings, Benjamin Cox; (5) consistency of the exercises with NCSS *Social Studies Curriculum Guidelines*, desirability and realistically satisfactory performance levels, June Chapin. Each of the investigators' reports was reviewed by the Steering Committee and, except for the first, by members of task review panels. The first, a delineation of the assessment model, is accompanied by two commissioned papers speaking to the model and prepared by Joseph Grannis and Michael Scriven. Those who worked on this project were not only competent but of varying professional roles in social studies education; attention was given to men-women ratios, minority groups and geographic areas. The project has aimed for thoughtful, honest and open points of view.

This volume reports on each of the five tasks of investigation, but with emphasis upon what has been less fully presented elsewhere. Readers will find another emphasis in a special issue of *Social Education* for May 1974, on National Assessment in citizenship and social studies.[1] The full NCSS report to National Assessment is available through the Educational Resources Information Center (See inside front cover). Since many people will find the publications of NAEP itself useful, a bibliography is included in this report as Appendix D. The chapters in this report examine the assessments from several stances. Both Taylor and Larkins treat the process of arriving at objectives, methods and procedures, but in differing contexts. Larkins looks at the objectives and Chapin's panel at the exercises with NCSS *Guidelines* as criteria. Hunkins's panel considered the validity of the exercises for the objectives, while Chapin's panel considered the worth of the exercises by the criteria of the NCSS *Guidelines*. Cox interprets findings from exercises categorized as social studies knowledge by the method of analysis, while Chapin's panel judges realistically satisfactory performance levels for individual exercises.

The project believed National Assessment, as well as other assessment programs, is better served by stating points of view that have both differences and commonalities. The full NCSS report and this volume do attempt, however, to consider basic questions: what is assessed, the extent to which the assessments can be counted on, what can be learned from the procedures themselves, what the findings mean and the extent to which they are useful.

So ambitious and potentially influential an undertaking as a national assessment of educational progress must from the outset make decisions on a host of issues. Some are highlighted here and elsewhere in this report.

A number of issues cluster around the matter of objectives. Who was to decide, first of all, what was to be assessed in citizenship and social studies — or any other areas? Teachers? Administrators? Those in state departments of education or those at the grass roots? Experts in social studies education? Scholars in relevant diciplines? Textbook authors and publishers? Minority groups? Students? Those who support custom in education, "what everybody knows it's always been," or those who support innovation, "the cutting edge?" Researchers on educational problems? Educational policymakers in legislatures and school boards?

Early in the enterprise and after consultation with several kinds of people, basic criteria for objectives were set. Objectives had to be (1) considered important by scholars, (2) accepted as an educational task by the school and (3) considered desirable by thoughtful lay citizens."[2] These criteria pointed to the kinds of people who were to decide, if not to the particulars of the process.

Then contracting agencies, in a sense, experts in assessment, searched the literature of the field out of which came a tentative list of objectives. Panels of those competent in social studies education or the fields of social science/history, teachers and other school persons, and thoughtful lay people concerned about education reviewed and revised until a set of objectives was formulated. How many kinds of people should be heard from? Were a sufficient number of persons from minority groups included? Were there too many professional, comparatively well-off people, and too few who could see education from the vantage of the poor? Were groups who needed information for policy making underrepresented? These are matters not readily resolved, especially within the necessary constraints of time and cost, nor closed off to future reconsideration. At any rate, it is noteworthy that thought from both the professional field and several sorts of people including laymen contributed to the formulation.

Such a process does not make for a theoretically clean and consistent set of objectives. On the one hand, it is foolish to disregard the contributions of scholarship. On the other hand, ours is a changing society and a pluralistic one. Not even scholars agree. No one official set of objectives exists, nor even one that draws wide allegiance, and probably least of all in the area of social studies education. Most of us like it that way. It can be argued that no single satisfactory set of social studies objectives is either possible or desirable. Against the merits of a theoretically consistent conception of social studies education must be balanced the need for a set of objectives that seem legitimate to many in society.

Almost from the start assessments in citizenship and social studies were separated. Much can be said in favor of two assessments rather than one focused on an area of critical yet controversial importance. Many will support the idea that citizenship is the responsibility of the school as a whole, not merely programs in social studies, or even that citizenship is as much the responnsibility of out-of-school institutions. But if social studies education can be thought of as emphasizing what is less likely to be learned informally in the culture at large, it can hardly be conceived as something without integral relation to individual lives and the requirements of society. Nor can citizenship be defensibly conceived as social participation without thought or knowledge. Neither is it sensible to think of citizenship as primarily political and social studies nonpolitical. Issues of distinction, overlap and emphasis are difficult to resolve.

Crucial also was the decision about what was to be assessed. National Assessment might have focused only upon some few basic skills, the three R's perhaps though they are far easier to name than identify. NAEP might have focused its efforts on assessing knowledge and knowledge only in some set of disciplines, or in learning areas commonly in school curricula. NAEP might have attempted to assess the outcomes of typical, or presumably "best," or "poorest" school programs, or, for that matter, out-of school educational institutions. The list of possibilities is long. The actual decision was for assessing a broad range of fields. As a consequence the assessment yields information about aspects of educational attainment in the population as a whole for which data have been sparse. Moreover, the decision throws the weight of the Assessment to a broad rather than narrow conception of educational attainment, especially important in social studies education. What is assessed exerts powerful influence on what schools see as important to teach and what students see as important to learn.

Another set of issues is embedded in the closely related area of exercises, expected to furnish evidence of attainment of the objectives, to be sure, but also to be significant in themselves. When objectives are translated into exercises, the chips are down.

Citizenship and social studies are inevitably touchy areas. To avoid what is controversial is in itself to take a position. To the credit of NAEP, it chose not to rule out the controversial. But how much and how sensitive? Review of exercises by panels of several sorts of people resulted in rewriting or even dropping a substantial proportion of exercises.[3] Social studies and citizenship were, indeed, more sensitive areas than some others. The issue here is much like that in the matter of objectives. What kind of balance can be had between the need for attending to the views of many diverse social groups, and the necessity for developing exercises legitimate in the eyes of many in the country at large?

One more dilemma appears in the matter of "right answers," especially for those exercises about complex problems yielding to no simple solution and/or depending upon points of view and attitudes. Are all positions taken to be considered proper responses if their holders support them with whatever reasons? Or, must some positions agree with predetermined proper responses, for example, support for the rights of the First Amendment? Or is some mixture appropriate? Blacks, native American Indians, to name two of a number of groups, experience significant differences in their social worlds from those of dominant groups. What should be considered proper responses — or proper exercises — for such groups?

Although National Assessment, unlike many school assessment programs, has not operated on a shoestring, it too must function within limits of financial support, time and the capabilities of the general field of assessment. NAEP too must make choices to what extent are self-reports in exercises justifiable substitutes for actual observation of "live" behavior? How much effort should be devoted to developing exercises assessing more complex, higher cognitive and affective behaviors? And, indeed, if such assessment requires much time from respondents, how much more time is feasible without throwing the baby out with the bath? Could special substudies do the job?

As does every assessment program, NAEP has had hard decisions in the construction of exercises. What they developed is a far cry, but a heartening one, from what many people have learned to think of as "tests." Exercises in social studies and citizenship frequently utilize paper and pencil but also rely on interviews, and even observations of discussion tasks. If young people were asked to respond to the familiar multiple-choice forms, they were also asked to view pictures; listen to songs; use maps, graphs, cards from a library card catalog and indexes; interact with each other in discussion groups; watch a film; and reply to interview questions. If not all, then many exercises are lively, innovative, readily related to the present social world and exemplary.

A last set of issues are those of interpreting the finding. Assessment in itself is neither evaluation nor explanation.

As succeeding rounds of assessments in citizenship and social studies are carried on, benchmark data will be available. Such comparisons can be made now in science. "On most exercises measuring science knowledge and skills, achievement declined at all three school ages assessed — 9, 13, and 17 years" from the first to the second assessments.[4] However, benchmark data can be had now for citizenship and social studies.

As a guide to interpretation NAEP has developed national percentages of successful performance for each exercise (and for some groups in the population and categories of exercises). Although illuminating, a few illustrations make it plain that these performance levels are not necessarily standards of what is "good," "adequate" or even "bad." When asked in the citizenship assessment whether a person on television or radio should be allowed to state any of three generally unpopular views, "statements that make some people angry," 3% at age 13, 17% at age 17 and 24% of young adults would allow all three statements and gave freedom of speech as the reason. Somewhat higher proportions would allow any one statement.[5] When asked to identify the meaning of monopoly in the social studies assessment, 51% of the 17-year-olds and 56% of the young adults could do so.[6] When asked in the social studies assessment to read a line graph identifying retail prices for eggs and apples over a period of time, 89% of 13-year-olds, 96% of 17-year-olds and 91% of young adults were able to do so.[7] But when the element of interpretation was included in an unreleased social studies exercise, fewer - 53% of those at age 13, 74% at age 17 and 69% of the young adults - could read and interpret a line graph.

National Assessment also compiles results by groups: age, region, sex, color, parental education, and size and type of community. For each of these groups differences from the national percentages of success are available, and comparisons among broad groups possible. Whether these differences are to be viewed with alarm or praise depends in part upon the size of the difference and the extent to which educational opportunity for all is accepted. What is more, if the national percentage of success is judged too low, a more successful group performance may still be inadequate.

The proportion who are able "to do" an exercise depends in part upon the difficulty of the exercise. Many standardized tests have been set up to distinguish the most able, typically able and least able, and exercises constructed accordingly. Although it might have been, such was not the purpose of National Assessment. Instead, it aimed to describe the educational achievements of the population at four age levels. Consequently, NAEP developed exercises in three approximately equal groups: for the least able, typically able and most able. Successful performance percentages must be read and results interpreted accordingly.

Of course, the Assessment might have followed still another path: hoping for mastery, claiming that everyone should be able to perform every exercise successfully. Such a course would have required either a set of exercises, within the reach of all, or a set of more-demanding exercises with a built-in and higher "failure" rate. The former would have failed to tap what the more able could do. The latter would have loaded the dice, emphasized not what has but what has not been attained. Much is to be said for "mastery." Perhaps we have all been too tolerant of "not getting it," moving on to something else before learning is achieved. The problem, however, comes down to agreeing on exactly what, specifically which, tasks every young person in this broad and diverse land should be able to perform.

The problems in settling on proper difficulty levels are again much like those in agreeing upon objectives. At any rate, performance levels over, let us say 90%, cannot simply and in themselves be judged as satisfactory, nor those below as unsatisfactory, pinpointing areas for improving educational endeavor.

National Assessment is almost inevitably caught between the frying pan and the fire. On the one hand, information from assessments can be better interpreted and used when tied to a school district or a social studies program. There, it seems at least, influencing conditions might be sorted out. Better yet, research questions and accompanying and flexible inquiry designs might point to influencing factors and implications for change. The meaning of data comes clearer when explanatory matters are related to outcomes, when questions posed give people a handle on the data. On the other hand, fears have been expressed from the outset that the Assessment might become a national testing program with all its consequent restraints. School districts or states or those engaged in some program or other — any identifiable persons or groups — could hardly be expected to enjoy or seek the glare of examination by some outside agency; sometimes it is enough to put up with those they can call their own. NAEP has intended to be neither a national testing program nor an examining agent. The thought of collecting data for identifiable institutions, programs or individuals on a national scale staggers the mind. All that means, however, that National Assessment does not furnish information directly to those making policy decisions for particular states or school districts or classrooms or social studies programs.

Moreover, a national assessment was conceived at a time when federal efforts in education were growing and the spotlight was on educational attainment in the country as a whole. By now efforts have in many respects shifted towards states and localities. They have, in turn, their own needs for evidence that a national assessment can satisfy only indirectly. Defensible assessment programs are costly, and shoddy or duplicate ones unjustifiable, all the more so in times of straightened economic circumstances. A mere collection of unrelated assessment data from some states here, some districts there could hardly allow for coherence in the whole or for information gathered in one place but useful in others.

In a sense, the strengths of National Assessment — and many others at state and local levels — have also been its weakness. What factors are to explain the findings? What produces what? For which policy questions are assessment data to be provided? NAEP is now addressing such problems by undertaking "special analysis activities requested by USOE to answer questions pertinent to federal policy decisions . . ." — for example, analysis of results for group combinations, such as race within region and community, to provide information on the matter of whether "the federal government should devise efforts to redress resource imbalance and for whom?"[8] The project has also commissioned a study of background factors affecting school achievement with an eye toward the feasibility of including some of these in Assessment studies.[9] Perhaps there are other means by which NAEP can include consumers of assessment findings in some ways like those of deciding upon objectives.

Still some points seem clear and a few are cited here. Typical performance on exercises in citizenship and social studies of school-age young people in the inner city is below the nation as a whole; typical performance of young people in well-off residential areas is above the rest of the nation.[10] Obviously enough large proportions of inner city young people are members of minority groups. Whatever can be said in support of arguments that the two assessments do not account sufficiently for the experiences of subcultural groups, it seems plain that attention must be paid. Social studies education and the multitude of conditions that influence it have to be better. The complexity of the problems does not justify sweeping them under the rug.

Social studies educators will do well to look at the results of specific exercises. Only 41% of 17-year-olds can respond properly to all five parts of an exercise on using a simple ballot. Nor do young adults pick this up once they become of voting age.[11] Surely those in social studies education ought to take steps to see that 17-year-olds in their own schools do better on a matter so vital. A number of race-related exercises show up a basic fund of decency among young people.[12] In the midst of conflict and change social studies education ought to find ways to capitalize upon it. Some 82% of 13-year-olds and 93% of 17-year-olds (and interestingly only 67% of young adults) believe that "teen-age students should help decide what courses will be offered in their school system."[13] Such expectations need to be accounted for in social studies curriculum planning.

What has been done and what has been found in the national assessments in citizenship and social studies is worth thoughtful consideration.

## *NOTES*

1. "NCSS Examines Aspects of the National Assessment of Educational Progress," *Social Education*, 38, No. 5 (May 1974).
2. See Appendix D, Reference 1.
3. See Appendix D, Reference 7, pp. 35-41.
4. Gaye Vandermyn, *National Assessment Achievements: Findings, Interpretations and Uses, Report No. 48* (Denver: Education Commission of the States, 1974), p. 7.
5. See Appendix D, Reference 5, pp. 34-35.
6. See Appendix D, Reference 8, p. 29.
7. See Appendix D, Reference 8, p. 15.
8. See Appendix D, Reference 10.
9. See Appendix D, Reference 11.
10. See Appendix D, Reference 3, pp. 79,81; Reference 8, pp. 19-22, 40-44, 55-59.
11. See Appendix D, Reference 12, p. 46.
12. See Appendix D, Reference 5, pp. 18-21; Reference 12, pp. 18-21.
13. See Appendix D, Reference 12, p. 2.

# CHAPTER 2

# POTENTIAL USES OF THE NATIONAL ASSESSMENT MODEL AT THE STATE LEVEL AND FOR ACCOUNTABILITY PURPOSES

Bob L. Taylor

University of Colorado

The basic task of this paper is to describe the model used by the National Assessment of Educational Progress (NAEP) gathering data and reporting on the citizenship area and to discuss the potential uses of the model for state assessment, curriculum development and accountability purposes. It is not within the scope of this paper to criticize the model with respect to technical flaws; hence, the model is described and discussed without reference to problems of design.

## The Model

The model is in the continuous process of being refined and improved; thus only its basic components are presented in Exhibit 1. A circular scheme is used in presenting it since, in reality, its actual application may be initiated with any one of the components. Also, in its actual application, there are continual interactions between and among the various components. While theoretically the process starts with the refinement of overall national goals into specific subject-matter, behavioral objectives and progresses in logical sequence through to the final utilization of information, in practice there is much greater freedom with respect to the utilization of the components.

The model for the citizenship assessment is presented here in outline form with a fairly detailed description of its components. As presented in Exhibit 1, there are seven basic components identified in the model: objectives development, exercise development, sampling plan, administration of exercises, scoring and analysis, reporting and dissemination and utilization of information. While many of the fine points of the model are not developed in the following outline, it is described in sufficient detail to give the reader a good understanding of how the data were collected and what implications might result from these data. The number of subtopics in the model and their distribution indicate that the major efforts of National Assessment have been with the first five components. The last three components have been areas of controversy; therefore, they have received less attention until recently.

## *Outline of the Assessment Model for Citizenship*

I. Objectives Development[2,3,4,5]

   A. The task of developing objectives in the field of citizenship was awarded to the American Institutes for Research (AIR) of Palo Alto, California. These criteria were used in examining the objectives:

      1. They were considered important by scholars.
      2. They were accepted as an educational task by the school.
      3. They were considered desirable by thoughtful lay citizens.

   Scholars reviewed the objectives for authenticity with respect to their subject fields, school people reviewed the objectives in terms of their actual emphasis in their schools, and laymen reviewed them in terms of their value in life.

   B. The AIR staff reviewed previous lists of citizenship objectives and reduced these to one comprehensive list of 20 objectives.

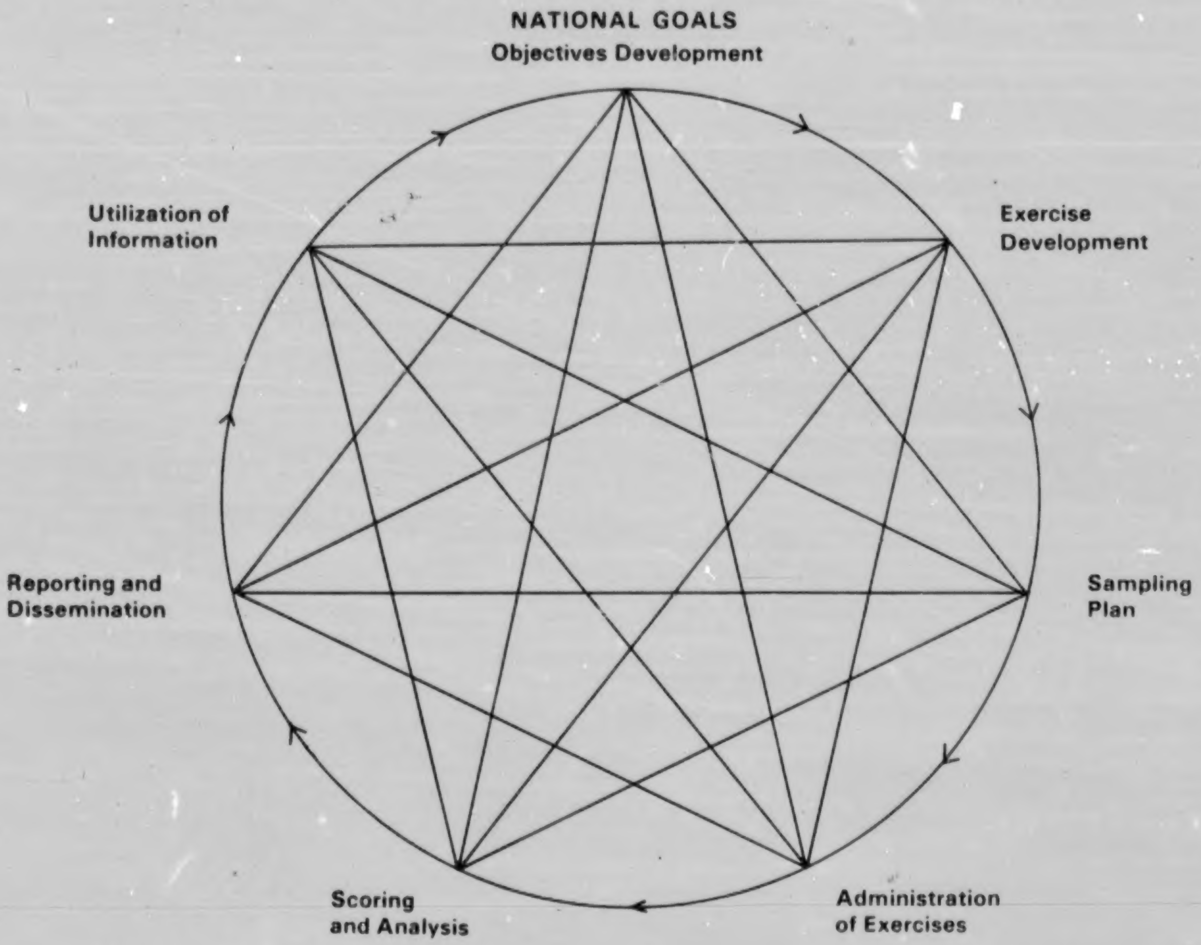
   C. Outstanding local teachers familiar with each target age level (9, 13, 17, adult), working with the AIR staff, broke down each general objective into the most germane behaviors deemed appropriate as goals for a given age level.

   D. A selected group of students and adults in each age level was asked by the AIR staff to recall and describe outstanding citizens of their acquaintance and specific incidents reflecting good and poor citizenship. These incidents and descriptions, about 1,000, were used to check the completeness of the initial list of objectives.

   E. The objectives were stated on three levels: general objectives, subobjectives and behavioral age illustrations or statements. The results were summarized for each age level.

   F. The revised list of objectives, broken down into important behaviors, was then studied for three days by a panel of national leaders in citizenship education and related social sciences.

EXHIBIT 1. Components of the Model


NATIONAL GOALS
Objectives Development
Exercise Development
Sampling Plan
Administration of Exercises
Scoring and Analysis
Reporting and Dissemination
Utilization of Information

G. A group of persons in various roles from selected California communities reviewed the objectives and made suggestions. These persons included public and private school administrators, counselors, teachers, a judge, a county planner, labor and business leaders and social scientists.

H. The objectives were then reviewed by panels of laymen. Eleven lay review panels representing four geographic areas of the country and three different community sizes were used. Each panel spent two days reviewing the objectives based on these two questions: "Is this something important for people to learn today?" and "Is this something I would like to have my children learn?"

II. Exercise Development[6, 7, 8]

A. The production of the exercises was initiated by AIR in 1966. The exercises were developed to cover all of the major objectives and to represent the selected content areas. Many exercises required the use of interview techniques, as well as the usual pencil-and-paper exercises. Self-report and group-task exercises were also used.

B. Because NAEP intends to describe what people in an age level can do, the exercises were written to reflect three difficulty levels — knowledge or skills common to almost all persons in an age level, skills or understandings of a typical member of an age level, and understandings or knowledge developed by the most able persons in an age level.

C. All exercises were developed to meet these criteria: content validity, clarity, functional exercise format, clustering exercises based on a single set of stimulus materials, directionality of response, difficulty level, content sampling and overlap between age levels. The exercises were direct measures of some pieces of knowledge, understandings, attitudes or skills that were mentioned in one or more of the objectives.

D. The exercises were reviewed by panels of lay persons for clarity, meaningfulness and invasion of privacy.

E. There was a tryout of the exercises involving representatives of groups in the actual assessment — regions, communities, races, sexes and age levels. Following the tryouts, the AIR staff and subject-matter specialists reviewed the tryout data and made needed revisions.

F. A committee of subject-matter specialists, measurement specialists and NAEP staff members rated the exercises to be included in the packages according to a set of criteria; the exercises were selected based on the ratings.

G. The selected exercises were reviewed by U.S. Office of Education personnel for any infringement of privacy on the part of the respondents or possible offensiveness.

H. Since there were about 160 minutes of testing time available for each age level in each learning area, the exercises used were only a small sample of the potential number of exercises. The exercises were assembled into administrative units (packages) for groups up to 12 persons.

III. Sampling Plan[9, 10]

A. The sampling plan was subcontracted to Research Triangle Institute, Raleigh, North Carolina. A multistage design, which was stratified by region, size of community and socioeconomic status, was used. This probability sample allowed researchers to collect data from a small sample of the population and to infer from that sample certain characteristics of the entire population.

B. The populations for assessment were all 9-year-olds, all 13-year-olds, all 17-year-olds and all young adults 26 through 35 years old in the 50 states plus the District of Columbia. The only exceptions were the exclusions of institutionalized individuals of these ages (those in hospitals, prisons) and others who could not be reached.

C. For ages 9 and 13, only a school sample was used, and for the 26-through-35 age level only household sample was used. For the 17-year-olds, both a school and a household sample were used.

D. The entire country was divided into population areas as follows: cities, counties exclusive of cities and pseudocounties — two or more counties were put together when the population of a single county was less than 16,000. Each population unit of 16,000 residents was assigned a number.

E. The country also was divided into four geographic regions: Northeast, Southeast, Central and West.

F. Each geographic region was divided into communities of four types: large cities of above 180,000 population, urban fringe, middle-sized cities between 25,000 to 180,000 population and small town/rural areas of under 25,000 population.

G. The 52 sampling units for each geographic area were spread across the four community types in a fashion proportional to their population in relation to the area population.

H. To insure comparable representation from each part of the coutry, an equal number of sampling units was selected from each geographic region — 52 from each of the four regions for a total of 208.

I. Sampling units were selected at random. This plan did not guarantee that all 50 states would be included in the sample. This was not a survey objective, but later the design was changed so each state was included in the sample.

J. In each sampling unit selected, all school buildings enrolling students of the sample ages (public, private and parochial) were identified.

K. The plan for schools was to select units of approximately 250 to 350 pupils for each age level and from at least two different buildings within each sampling unit for each age level.

L. Each cooperating building principal provided a list of names of students in the building from the specific age levels. This list was used for the final random selection of students to take the assessment exercises from that building.

M. Information about the areas was obtained from U.S. Census data. In order to insure reliable information for lower-socioeconomic groups, these groups were oversampled. There was a disproportionate number of schools from lower-socioeconomic areas included. In the overall results, the data from the lower-socioeconomic areas were given the percentage value in which they occurred in the total population.

N. From each of the 208 geographical samples, 100 adults, ages 26 through 35, were randomly selected using the following procedures. Each of the 208 geographic samples was divided into equal secondary sampling units. Then 10 secondary sampling units were randomly selected from the total 208 samples. Interviewers then personally contacted the people in the chosen secondary sampling units of the 26-through-35 age level and out-of-school 17-year-olds. These persons were asked to participate in the assessment.

O. Individuals were classified as black, white and other on the basis of information provided by the school or by observation. Results were given for black and white only. The number of individuals classified as other was too small to produce reliable results.

IV. Administration of Exercises(1, 2)

A. Administration of the exercises was subcontracted to Research Triangle Institute in the East and to Measurement Research Center of Westinghouse Learning Corporation, Iowa City, Iowa, in the West. Cooperation of schools was obtained by first contacting officials at the state and then at school district levels. There was above 90 percent cooperation by schools. Adults and out-of-school 17-year-olds were contacted by a personal door-to-door household canvass. Each out-of-school participant was contacted individually. The right of each to refuse to cooperate was respected.

B. A full-time trained staff of 27 district supervisors managed the fieldwork. They were assigned to different geographical areas of the United States. They contacted schools and recruited and trained local teachers to help in the administration of the exercises in schools; they recruited and trained other available persons for the out-of-school administration.

C. In the schools, students from a single age level from different classes were brought together in a room for exercise administration. Group size was at least 8, and usually 12, students.

D. The exercises were organized in packages that contained exercises from two or three different learning areas at a single age level. No one person took all the exercises in his age level. Age levels were assessed at different times of the year.

E. In packages administered to groups, taped directions and taped readings of the exercises were used in addition to printed packages. This was done to establish consistency in timing and administration plus to provide for nonreaders.

F. Several packages at ages 9, 13 and 17 consisted of exercises that were given by exercise administrators to one individual at a time. The administration of all the packages for the adult assessment was done by interviews.

G. Each package required about 50 minutes of administrative time. Each person took only one package with the exception of the out-of-school 17-year-olds, who were asked to take four or five packages each since they were the most difficult and expensive group to locate.

H. Students' names were confidential and did not appear on any packages. The name roster was kept at the building level and used only in the organization of the in-school sampling.

V. Scoring and Analysis[illegible]

A. The scoring and analysis of the exercises were subcontracted to Measurement Research Center of Westinghouse Learning Corporation, Iowa City, Iowa.

B. The multiple-choice exercises were scored and recorded routinely by machine.

C. The open-ended exercises were scored by trained professionals using a key of acceptable and unacceptable achievements in terms of the objectives.

D. Results were reported for each goal. Also, the results were reported both as (1) the percentage of any group of respondents making the desired responses to an exercise and (2) the difference between the percentage of a group making the desired responses and the corresponding national percentage.

E. In the assessment, there was a lack of proportionality among characteristics used in the comparison of groups, such as color, sex, parental education. A statistical procedure, balancing, was used to correct for this problem in the comparative analysis of the data. Balancing is a procedure to examine the performance of groups classified on one characteristic adjusting for the fact that these groups differ on a specified set of other characteristics.

VI. Reporting and Dissemination[illegible]

A. The reporting of results was directed to subject-matter specialists, professional educators and informed laymen. Multiple reports were developed to serve these different audiences.

B. Approximately 40 percent of the exercises were reported at the end of each assessment year. Not all exercises were reported since they were to be used over again in future assessments in order to measure change.

C. The exercises released for publication were selected to be representative of all exercises administered as well as the results received on the assessment.

D. Reporting was done by age levels. Since some exercises were used with different age levels, there were often comparable data across two or more age levels.

E. Reporting was also done by groups within the categories of regions, community types, sex, socioeconomic status and color.

F. Reports were printed with a short description of the exercises, the national percentage of success and group differences from the national percentage of success for each exercise and without interpretation of results.

G. Both observed and balanced results for all exercises were reported by groups. The effects of balancing on measured characteristics such as sex and region were included in the report.

H. There were no scores reported for individuals. No single individual took more than one twelfth of the exercises, and no individual took a package that sampled a single learning area.

I. Results were reported through the media: written word, radio, television, films and personal reports.

VII. Utilization of Information[18, 19]

A. The results provided potential information for education decision making. For example, considering the somewhat lower performance of the Southeast region on the citizenship results, school boards in that region might decide to put greater stress in their school programs on citizenship skills, understandings and attitudes.[20]

B. The results raised many questions that may lead to other investigations. For example, in making comparisons of all citizenship results combined, it was found that the extreme affluent suburbs showed substantial median advantages at all ages and that the extreme rural and extreme inner city showed substantial deficits at all ages. These discrepancies in performance need causal studies conducted from the perspectives of different disciplines such as political science, sociology, economics and education.[21] Comparisons might lead to other studies: a sample of 5-year-olds might be assessed for a basis of comparison; results might be broken down by states; or new learning areas might be investigated.

C. The results of several cycles should provide evidence of the change in knowledge, skills, understandings and attitudes in the age levels as they relate to education objectives.

D. School administrators can make comparisons between groups and may improve student performance from the information gained in this manner.

From this review of the model, it is evident that the NAEP staff has put a great deal of effort and know-how into the design, plus the development, of each of the components.

In summary, it can be pointed out that the model reflects some important choices on the part of the NAEP staff. The decision was made to assess a broad range of objectives in each of the 10 learning areas. Certainly, it would have been much easier and cheaper to have concentrated on a narrow set of objectives. Also, the effort was successfully made to include the higher cognitive levels in the assessment exercises and to deal with the affective domain. In learning areas like citizenship and social studies, the usual standardized test concentrates on factual knowledge, which is of a less controversial nature and easier to assess. National Assessment should be commended for its bolder, more comprehensive approach to the task, which searches for more significant kinds of data.

## Use of the Model at State Level

National Assessment is a census-like study to collect information concerning the educational attainments of Americans. In planning for the collection of this census-like data, the model, which was presented in the previous section, was developed. A number of states have found adaptations of the model useful in conducting state assessments in which desirable learning outcomes are identified and the status of learners with respect to these outcomes is determined.

State assessment is a rapidly developing movement. At this writing, all of the states have assessment activities either in operation, in a developmental process or in a planning stage.[22] While the statewide assessment programs have many similarities, they break down into two basic types of programs on the question, "Who gets to use the results?" The divisions are those states for which data are collected for decision making by state agencies and those states for which data are collected for decision making by teachers and administrators. State programs for which the emphasis is on collecting information for state-level decision making are: Arizona, California, Colorado, Connecticut, District of Columbia, Florida, Maine, Massachusetts, Michigan, Nevada, New Jersey, New York, North Carolina, Rhode Island, South Carolina, Tennessee and Texas. Programs for which the emphasis is on collecting information for local-level decision making are: Alabama, Arkansas, Delaware, Hawaii, Idaho, Iowa, Kentucky, Mississippi, New Hampshire, New York, North Dakota, New Mexico and Pennsylvania.[23]

Beers and Campbell report that a number of characteristics are appearing in these assessment programs. In about a third of the states, the programs were mandated by the state legislatures, and the results of the assessments are to be reported back to the state legislatures. In a few of the states, the data are to be used for Planning, Programming and Budgeting Systems (PPBS). In about half of the states where the assessment data are being used to make state-level decisions, state and federal funds will be allocated based on the results. Participation in assessment is required by law in about a fifth of the states. In many states where the assessment data are being used to make state-level decisions, samples rather than all students are being assessed, while at the local level, all students in the target populations are being assessed. Criterion-referenced instruments are very common with the states where the data are being used for state-level decisions, but the states collecting information for local decision making are favoring norm-referenced instruments. Finally, no dominant funding pattern has evolved in either of the two groups of states.[24]

### *State Adaptations of the Model*

In its assessment of citizenship education, Maine made an extensive application of the NAEP model and carefully duplicated it to collect comparable data at the state level.[25] Maine's first cycle of the 10 learning areas of the Assessment is to be completed by scheduling two of these areas each year for five years. Citizenship and writing were the first learning areas to be assessed.

Based on the results of a previous study of objectives for education in Maine, two review committees decided to accept the National Assessment objectives as being closely related to the Maine objectives. Maine selected the 17-year-old population of in-school students for its first assessment. A sample of 2,000 17-year-old students was used to represent the approximately 17,000 17-year-old students in the state. The state was divided into four geographical regions. As in National Assessment, school buildings were randomly selected from the geographic regions, and students were then randomly selected from the buildings. Packages were developed with exercises taken from the two learning areas. The available, released exercises from NAEP were carefully examined to see if they reflected objectives valid for Maine and to see if some could be modified, where needed, to be administered in group sessions using the paced-tape method while still retaining a high degree of comparability to the National Assessment individually administered exercises. The packages were made up of 23 citizenship and 7 writing exercises, plus a 23-item student questionnaire. The exercise format was kept virtually identical to the one used in National Assessment. Trained administrators were sent out to administer the exercises, and the exercises were scored according to NAEP procedures. On data reporting and analysis, there was the census-like reporting of the performance of the Maine students plus comparisons of the Maine results with appropriate National Assessment data.

In summary, the Maine assessment duplicated NAEP procedure as nearly as possible. With minor exceptions, the same objectives were used for citizenship. The same sampling design was used with adaptations to a smaller geographic area and population. The exercises were, for the most part, taken from those released by National Assessment, and they were organized into packages similar to those used by National Assessment. The administration and scoring of the exercises were conducted in the same manner as NAEP. Since Maine used the same private contractors as National Assessment, the duplication was complete wherever possible. The reporting and data analysis were similar, and the data did provide the opportunity to compare Maine's results with those of National Assessment.

Here, the model was very carefully duplicated at the state level. The big question that comes to mind after studying the Maine citizenship report is, "Aren't the National Assessment data being treated here as some kind of a national norm against which the performances of 17-year-old students in Maine were being compared?" Of course, this use of Assessment data had been questioned from the start of the proposal for an assessment at the national level. Now Maine has provided the opportunity to study the effects of this use of the data on the education system of a state.

Another state that carefully followed the model was Connecticut,[26] where an assessment was first conducted in reading. To permit comparisons, the Connecticut program used available instruments and applicable procedures developed by NAEP, but adapted to the requirements of the local situation. Connecticut's reading objectives were matched to the reading objectives of National Assessment. Approximately 220 reading exercises from NAEP were used in producing the packages used in the Connecticut assessment. Exercises were selected to represent all of Connecticut's reading objectives. The age levels assessed were 9, 13 and 17. As with the National Assessment packages, tape-recorded instructions were used. The sampling design was a multistaged one duplicating with few exceptions the National Assessment design. As with National Assessment, a group of administrators for the packages was recruited and trained.

The Connecticut assessment was another example of careful duplication of the NAEP model — even using the same objectives and exercises. Again, National Assessment results were used as norms to which the state results were compared.

The Texas Needs Assessment used the model for the development of their assessment in mathematics at the sixth-grade level.[27] However, while using ideas from the model, the Texas people broke with it in a number of places. They were concerned that the assessment would yield information that would be useful to teachers in their classroom instruction of students. From a pilot study, it was decided to use a criterion-referenced reading test and to work with grade levels instead of age levels of students. They worked with the sixth grade, and the tests were administered by the staff of each school that participated in the assessment. The objectives were chosen from the major skill areas treated in the state-adopted textbooks. Regional location and community size were taken into consideration in selecting the sample. Approximately 10 percent of the Texas schools teaching at the sixth-grade level administered tests, and approximately 10 percent of the pupils in the sixth grade were included in the sample. Reports were given to teachers on the performance of their individual students. Also, there was a school report on the performance of the students for each school and a report on each of the classes in the school. Comparisons were made on the basis of sex, race and size of community.

The Colorado Needs Assessment, while using the model, made an even greater break with it.[28] Its objectives were based on a state study of education goals, and these goals were restated in terms of performance objectives. Following the model, objective-referenced exercises were written. A sampling design was used and the student responses were analyzed. In this assessment, classroom teachers were involved in the writing and refinement of the behavioral objectives. Objective-referenced exercises were written for nine learning areas. The exercises were administered to a sample of 30,000 Colorado students. A stratified, random-sampling procedure was used to select a sample of school districts of the state. Then schools were selected at random from the districts chosen. Finally, classes in school buildings were randomly chosen for testing. The samples were representative of all Colorado students in grades 3, 6, 9 and 12. A group of proctors was hired and trained to administer the exercises, and the exercises were scored by computer. The data were analyzed on a statewide and district basis, and the results were broken down by groups, e.g., boys, girls, urban, rural.

As pointed out earlier, states are rapidly moving into the assessment field. Some are reproducing the NAEP model at the state level, and others are developing variations of the model. The cruder efforts have resulted in endless pages of raw percentages without any explanation of the results. Based on a survey of state assessment programs, Beers and Campbell identified several of the problems that are common to these state programs.[29] Naturally, a shortage of money and staff was the most frequently mentioned problem, for it is a fact that many states have moved into this area without providing adequate funds for a realistic assessment program. Also, teacher resistance to assessment and negative public attitude toward outside testing were problems mentioned. Test results have been misused in the past; teachers have been fired, for example, on the basis of incorrect interpretation of test results. Also, test developers have been guilty of violating the privacy of students through questions that transgressed the examinee's human and legal rights. A third problem area has been with the utilization and dissemination of resaults. Some school officials do not understand the results. In some situations, there has been hostility to the results. Some officials have ignored results in making decisions. Finally, results have frequently not reached the right people in a usable form.

### *Adaptation of the Model*

In the above discussion, it is evident that there will be as many adaptations of the model as there are state units conducting assessments. Probably there is no specific assessment model that is the best; hence, there is no model that should be applied without modification in any and all situations. Nevertheless, there are principles of good assessment that should be applied in developing or adapting a model for state assessment purposes. Listed below are some characteristics that should be found in a good assessment program.[10]

1. The program has clearly defined goals that apply to a particular audience or audiences.
2. The program has a realistic number of goals that are attainable under the existing assessing conditions.
3. The program has established priorities among its goals and places its major efforts on its major goals.
4. The program has been designed to gather information considered to be important in education.
5. The program has specific objectives, which it is striving to attain.
6. The program has been designed to provide results at a usable level of accuracy.
7. The program has used data-gathering instruments that measure the objectives of the assessment.
8. The program has collected data in such a manner as to introduce a minimum of error in the results.
9. The program has scored and processed data in an accurate manner.
10. The program has used analytic techniques that provide the data breakdowns needed by decisionmakers.
11. The program has reported results in a manner usable by its audience.
12. The program has provided help in the interpretation of results and assistance in their implementation.
13. The program has provided for the active involvement of groups of persons from all of the major audiences for the assessment results.

## Implications of the Model for Curriculum Development

Of course, a major, potential outcome of National Assessment and the model was providing new, accurate data with regard to curriculum problems. Curriculum decisionmakers are furnished data that have not been available to them before this. Because of this new information, they should gain new insights into their problems; hopefully, innovative approaches to the solution of these problems will result.[11]

### *Implications of the Model*

The assessment model has potential for promoting curriculum development. This is especially true when it is applied to state situations in the manner used in Colorado. In this situation, objectives were developed that specifically applied to the local situation. The statement of well-written objectives in behavioral terms may sharpen the purposes of instruction. Through the experience of writing behavioral objectives, the curriculum worker gains a much clearer perception of his task; hence, this practice may have a beneficial impact on curriculum work. On the other hand, the use of behavioral objectives has not always been a positive influence. The objectives may zero in on easily defined behaviors that lack scope and significance. They may produce tunnel vision and put stress on the inconsequential and trivial. In an effort to be specific and to define the exact behaviors desired, the larger perspective may be lost.

Again, the development of exercises from the identified behavioral objectives may have a positive influence on curriculum. The kind of innovative exercises that have been developed by National Assessment may have a very positive influence on what is being taught and how it is being taught. Teachers, both in reviewing exercises that have been used in National Assessment and in writing exercises for local assessments, may be influenced in their selection of both content and methods by their knowledge of these assessment exercises. Material not relevant to the objectives of the course may be dropped, and methodologies promoting the kind of skills needed in the assessment exercises may be introduced.

On the other hand, the results may be less desirable. If in state situations the dictates of finances or the lack of leadership result in the use of poorly wriitten, machine-scored, merely multiple-choice exercises, the results may be very negative. Teachers may feel pressured to stress rote learning of facts in order to prepare their students for poorly written examinations. Hence, poorly written exercises may keep irrelevant material in the curriculum and limit curriculum innovation and development. The quality of the exercises written and released will have an impact on curriculum development.

Sampling procedures may give insight into the status of knowledge, understandings, skills and attitudes of students in a particular target population. Findings from these procedures can promote curriculum improvement and innovation. Problem areas in the curriculum may be identified. Results of the National Assessment in citizenship have identified some problem areas. On an exercise dealing with freedom of speech, a large percentage of 13 and 17-year-olds indicated that they would not allow sample controversial statements to be made on radio or television.[12] This kind of response indicated a lack of understanding or valuing of the Constitutional right of freedom to express controversial or unpopular opinions.

On the other hand, there are potential difficulties with assessment data that represent national levels of performance. Even though the data were not collected with this intention and were reported in census-like form, the results of National Assessment are being treated like national norms. Several states have conducted their own assessments duplicating the NAEP model so that they can make direct comparisons between their state results and the various national, regional and group results. There is the potential of great mischief in this approach, for it may lead to unfair comparisons between groups, states and regions. In the assessment reports of some states, tables of percentages have been presented without any interpretation or explanation.

Some school systems have been presented in a very bad way without any reference being made to the kinds of variables involved in the different learning situations. Such variables as per-pupil expenditures, educational level of parents and motivation of pupils do have an impact on the learning situation. These and other variables should not be ignored in interpreting assessment results.

It is not suggested here that assessments should not be conducted because of the potential misuses of the data. Instead, it is suggested that those engaged in assessment at national, state and local levels have the responsibility to report data in proper perspective and to aid those using the data to make correct interpretations of it. We need these kinds of information for decision making, but if the data are misused or misinterpreted, then the decisions based on them may not be good ones.

### *Impact on Curriculum*

National Assessment is providing data on which decisions can be made. The reports on science and citizenship have resulted in strong recommendations for curriculum changes in these learning areas.[13] As data are gathered at the state level through the use of the model and its adaptations, specific suggestions for changes may be made. For example, Texas has designed its state assessment so that there is direct feedback at the classroom level. This may have a strong and immediate impact on these classrooms, either good or bad depending on what interpretations and recommendations are made.

Still other problems should be pointed out concerning curriculum decision making and national and state assessment results. While the National Assessment process for identifying objectives provides for broad-based participation in the decision-making groups, objectives still are selected that neglect sizable, sub-cultural groups in our nation. Even some of the very general objectives selected may not apply to these groups, and thusd they are not valid for some situations.[14] To illustrate the potential difficulty with objectives and groups, consider Objective V of social studies: "Have a reasoned commitment to the values that sustain a free society." Under it is subpoint B, "Believe in the rule of law and can justify their belief." For blacks in the core city, this objective may not carry value. Their experiences may have been that the law is used against them, and they constantly need to be looking for new ways to work around the law.[15] While this is perhaps an extreme example, it does make the point that in as diverse a population as we have in the United States, there will be many groups for whom the broad general objectives do not carry the same meaning as for other members of the nation.

There is also the question of what the components of a general education are or should be. Selecting common objectives for a learning area such as science and writing exercises for these objectives provide one definition of what students of a certain age level are expected to know about science. Since this establishes what comprises general education in science, it has a definite limiting impact on a student's freedom of choice in deciding what he wants to learn. He is being dictated to concerning what he should take in general education. This situation always has existed in education, but it puts the assessment movement in the camp of the conservatives in the current controversy with respect to free choice and unlimited electives for students. Here the assessment movement is counter to the humanizing movement in American education.[16] It is promoting a closed rather than an open approach to curriculum.

The very identification of a learning area for National Assessment will have an impact on the fortunes of that learning area. The "chosen" learning areas are more likely to be supported financially and retained in the curriculum than those not selected by National Assessment. Those learning areas that are not included in the "magic" 10 may find that they are second class as far as school boards are concerned. If the students of the district have made a poor showing on the state assessment, funds may be shifted to the support of those subjects where the low scores were identified. The learning areas that are not getting this public exposure may find their financial support reduced. While the potential for better education decision making is here, there is also the potential for poorer education decision making because of the impact of public exposure of the assessment results on the decisionmakers. The foreign languages are not included in the 10 learning areas assessed; hence, language departments will not be able to cite assessment data as evidence that there is need for greater support of their programs. They have been put into a poorer bargaining position by this omission.

Moreover, there is potential for "shortcut" assessment schemes by publishers, although as far as this writer knows, no assessment instrument of this type has been produced to date. Why go to all of the work for an expensive local assessment effort when you can buy a commercial assessment package, which is based on the released NAEP exercises and which will provide the school district with results that may be compared to NAEP results (norms)? The exercises may be given and scored by the local teachers, and the results may be compared with the National Assessment results included with the commercial assessment package. If the project is handled "right," a "live-wire" superintendent can demonstrate at a modest cost that his district is outscoring the National Assessment results (norms).

The idea of a commercial assessment package based on released NAEP exercises has positive potential. With honesty and careful application, this approach could be of considerable value to school districts. First, the objectives being assessed by the exercises included in the instrument need to be carefully identified. Then if the school district finds that these objectives fit their own objectives adequately, the exercises may be used with confidence. Second, there is no reason why teachers should not be able to conduct and score the exercises accurately. A tape-paced administration could be used, and the directions for scoring could be written in such a way that teachers could follow them with good results. District-collected data about the performance of students on NAEP exercises could be valuable data, if collected at a modest cost. There are a good many "ifs" and pitfalls in this proposal, but it is true that accurate, valid data could be collected in this fashion. Undoubtedly, some districts will find this to be a workable plan.

Cost is one of the big problems. It took a great deal of money to develop National Assessment's sophisticated program. Currently, the data being provided by the Assessment are of high quality, and the results have great promise for promoting education improvements. Nevertheless, hundreds of local districts are considering their own assessments, and the question of how to reduce the cost is becoming important. Should they develop their own assessment instruments, or should they use a commercial version of the model? For many districts, it is not possible to develop their own assessment programs; hence, for many reasons, the second alternative will be used. Over the coming years, the commercial assessment package will probably become a reality. The only other alternative open to poorly financed districts is to resist the pressure to become involved in local assessment.

Another potential outcome of an assessment program is the spin-off of research projects designed to investigate questions of causality raised by assessment results. It has been illustrated several times in this chapter that further investigation is needed to interpret results more productively. Any number of potential doctoral studies in NAEP data need to be made before the results may be confidently used in curriculum work.

A final interesting prospect in this entire assessment business is the increased pace of change in our society within recent decades. Those of us who have been involved in education decisions over the last couple of decades know that catching the direction of things is frequently more important than trying to make careful, data-based decisions. By the time we have collected and analyzed our data base, society and technology have gone off and left us. It is hard to fault the soundness of the data-collection design of National Assessment, but there is a five-year time lapse from start to finish of a cycle plus the time lag of scoring and data analysis. The question is whether this is the best process for education decision making in the last quarter of this century. Are changes in our society coming so fast that long before the data base is established the data are no longer relevant for the decisions for which they were gathered?

## Use of the Model for Accountability Purposes

Assessment is not the same thing as accountability, for accountability places greater emphasis on value judgment than assessment does.[37] Accountability is concerned with the badness or goodness of something. Education assessment is aimed at improving decision making by collecting information concerned with the outcomes of education. Accountability has varying meanings depending on who is writing about it.

A number of approaches have been proposed to make schools more accountable, such as a systems approach, management by objectives, education-program auditing, a planning-programming-budgeting system, performance contracting, voucher plans and alternative forms of education.[38] A widely accepted interpretation is that accountability is to determine whether the teacher, who is assigned the task of educating a group of students, is performing that task.[39] On the other hand, a broader interpretation of accountability is that it is a process for determining whether the program of a school, district or state is producing the student achievement expected with regard to the objectives of the program.[40] The first of these interpretations puts the responsibility for individual pupil achievement on the teacher. The second places the responsibility for the output of a program of instruction on the school, district or state school system.

The first interpretation of accountability has been widely debated and has gained the opposition of many groups including the teachers' organizations. A model of this approach to accountability follows.

I. Behavioral Objectives
II. Stated Evaluation Criteria Related to the Objectives
III. Scheduled Materials, Learning Activities, Equipment, Etc.
IV. Teaching Activities
V. Evaluation of Student Performance Based on Behavioral Objectives

If the use of this model stresses the output of the teacher in terms of pupil achievement, it does not provide for the multitude of variables that are found in any teaching situation. The primary emphasis becomes "Are the children learning what the teacher was hired to teach them?" No allowances are made for the many variables such as pupil ability, parental education or wealth of the school district, all of which and more may influence the success of the teaching in a given learning situation.

Under this approach to accountability, complete data must be collected for each individual in the population. It is necessary to identify the performance of specific individuals with respect to stated objectives since the responsibility for the individual student's performance is to be assigned to the instructor charged with teaching for these objectives. Here, it would seem, accountability is taking us back to yearly, mass testing with which some of us are only too familiar. Test administration procedures, such as NAEP's tapepaced method, could be put to work, although, because of financial considerations, teachers would probably administer these mass tests with many of the problems in the results that have emerged in the past.

The second interpretation of accountability presented here also sees where students are or are not achieving, but it is concerned with what the strengths and weaknesses of a program are and how they relate to student achievements. The NAEP model may easily be adapted for this type of accountability use. In assessment, the objectives are identified, the achievement level with respect to these objectives in a population is determined through a sampling assessment procedure, and the results are reported in terms of what percentages of the population are achieving the objectives.

The National Assessment model, as it was originally designed, did not provide usable data for determining who was responsible for the individual student either achieving or not achieving the stated objectives. As designed, the model provided information about the achievement on the stated objectives of a population or groups of that population; hence, it was not possible to identify the results for individuals in these population groups. Neither, was it possible to establish what individual teacher was responsible for the students having either achieved or not achieved the stated objectives.

Now let us examine the National Assessment model for its application to the evaluation of an entire curriculum for accountability purposes. The model components identified in this paper are: objectives development, exercise development, sampling plan, administration of exercises, scoring and analysis, reporting and dissemination, and utilization. Under accountability, objectives may be developed with input from a number of sources, including the patrons who are paying the bill for education, or an already established set of objectives may be used.[42] Since this is the evaluation of a given curriculum, the objectives should apply to that program if the results are to be valid. Who should select these objectives is an issue in accountability. Shouldn't there be input here from the teachers who are presenting the program?

While there are some problems in how objectives are to be selected, the components of the NAEP model are all functional in this adaptation for accountability purposes. Measurement specialists contend that objective-referenced exercises written for specific learning objectives are better for use in accountability evaluation than the norm-referenced exercises commonly used in standardized tests.[43] Therefore, released objective-referenced, National Assessment exercises may be used for accountability purposes where their objectives are valid for the learning situation that is being evaluated.

In the National Assessment design, the samples were drawn in such a fashion that they represented the population of 9, 13, 17 and 26-35-year-olds and groups from these populations. Where the model is being used to evaluate the performance of a given population, sampling procedures may be used as they were developed for the Assessment model. The same professional care that is taken in exercise administration of the NAEP model would be desirable in the application of the model for accountability purposes. Scoring and reporting are the same for assessment and accountability, and the same percentage kind of reporting can be used in both situations.

Where accountability is being applied to a total organization such as a school, a district, or a state, the National Assessment model may be used with little modification. It was designed to establish accurately the level of performance on a given set of objectives in a population, and it can do this for accountability purposes as well as assessment purposes. Likewise, it can assess groups of the population so that specific strengths or weaknesses in their performance can be identified.

There are many value questions related to accountability, and while it is not the task of this report to discuss them, a number are cited here. Will tight state accountability structures severely limit creativity and innovation in the schools? As indicated in some of the previous comments, doesn't accountability have the potential of becoming a stultifying state-testing program? Will the single-minded pursuit of achieving performance objectives lead to the abuse of children? Couldn't a disproportionate amount of time be expended on the defining of objectives without a commensurate increase in learning? Could not objectives established by politically oriented groups set schools on a course of indoctrination? Currently, in education there is a humanist-behaviorist conflict. Doesn't the accountability movement support a kind of techno-urban fascism?[44] Could not the accountability movement lead to the teaching of easily quantifiable material and discourage the inclusion of material that is difficult to quantify? Doesn't accountability promote a closed rather than an open education system? Doesn't accountability present obstacles to the continued development of freedom and autonomy for teachers? This is not an exhaustive list, but these are all disturbing questions. Furthermore, there is considerable evidence in the literature that supports the seriousness of the issues raised by these questions.[45, 46, 47, 48, 49]

## Summary

Several summary statements can be made. The National Assessment model is no doubt the best and most comprehensive procedure designed for collecting data for these purposes. The model can and has been successfully adapted for use at the state level; and as states become more active in assessment, many adaptations of it will be made to fit local needs. Finally, where the concern is with the evaluation of group performance, the model may be used for accountability purposes. With the current rapid development of state assessment and accountability programs, it is expected that the model will be widely used for these purposes in the coming years.

### *NOTES*

1. See Appendix D, Reference 16.
2. See Appendix D, Reference 1.
3. See Appendix D, Reference 2.
4. Vincent N. Campbell and Daryl G. Nichols, "National Assessment of Citizenship Education," *Social Education*, 33 (June 1969), pp. 279-81.
5. See Appendix D, Reference 5.
6. See Appendix D, Reference 16.
7. See Appendix D, Reference 9.
8. See Appendix D, Reference 7.
9. See Appendix D, Reference 13.
10. See Appendix D, Reference 1.

11. See Appendix D, Reference 16.
12. See Appendix D, Reference 9.
13. See Appendix D, Reference 16.
14. See Appendix D, Reference 9.
15. See Appendix D, Reference 16.
16. See Appendix D, Reference 5.
17. See Appendix D, Reference 9.
18. See Appendix D, Reference 16.
19. Larry E. Conway, "Some Implications of the National Assessment Model and Data for State and Local Education" (paper presented at the 1973 Annual Meeting of the AERA, New Orleans, February 26, 1973).
20. See Appendix D, Reference 4.
21. See Appendix D, Reference 3.
22. Joan S. Beers and Paul B. Campbell, "Statewide Educational Assessment," *State Educational Assessment Programs, 1973 Revision* (Princeton: Educational Testing Service, 1973), p. 1.
23. Ibid.
24. Ibid., pp. 2-3.
25. *Maine Assessment of Educational Progress: Methodology, Report 5* (Augusta: Department of Educational and Cultural Services, 1972).
26. *Report on the Assessment of Reading Skills of Connecticut Public School Students* (Philadelphia: Institute for the Study of Inquiring Systems, and Hartford: Connecticut Department of Education, 1972).
27. *Sixth-Grade Mathematics: A Needs Assessment Report* (Austin: Texas Education Agency, 1972).
28. John W. Helper, *An Assessment of Learner Needs in Colorado* (Denver: Colorado Department of Education, 1972).
29. Joan S. Beers and Paul B. Campbell, "Statewide Educational Assessment," *State Educational Assessment Programs, 1973 Revision* (Princeton: Educational Testing Service, 1973), p. 3.
30. Frank B. Womer, *Developing a Large Scale Assessment Program* (Denver: Cooperative Accountability Project, 1973), p. 89.
31. Marjorie M. Mastie and Frank B. Womer, "How Will National Assessment Change American Education?" *Phi Delta Kappan*, 53 (October 1971), pp. 118-20.
32. See Appendix D, Reference 5, p. 34
33. Charles H. Harrison, "Are We Educating for Tomorrow?" *Scholastic Teacher*, September 21, 1970, pp. 16-19.
34. J. Wayne Wrightstone et al., "Accountability in Education and Measurement Problems," *Test Service Notebook 33* (New York: Harcourt, Brace, Javanovich, Inc.), p. 3.
35. See Appendix D, Reference 14.
36. Stephen P. Hencley, "Impediments to Accountability," *Administrator's Notebook*, Vol. XX (Chicago: University of Chicago, December 1971).
37. Frank B. Womer, *Developing a Large Scale Assessment Program* (Denver: Cooperative Accountability Project, 1973), p. 3.
38. Lesley H. Browder Jr., *An Administrator's Handbook on Educational Accountability* (Arlington, Va.: American Association of School Administrators, 1973), pp. 14-21.
39. J. Wayne Wrightstone et al., "Accountability in Education and Measurement Problems," *Test Service Notebook 33* (New York: Harcourt, Brace, Javanovich, Inc.), p. 1.
40. Lee J. Cronbach, "Evaluation for Course Improvement," *Readings in Measurement and Evaluation*, ed. Norman E. Gronlund (New York: Macmillan Company, 1968), p. 41.
41. Lesley H. Browder Jr., *An Administrator's Handbook on Educational Accountability* (Arlington, Va.: American Association of School Administrators, 1973), p. 50.
42. Irving Morrissett, "Accountability, Needs Assessment, and Social Studies," *Social Education*, 37 (April 1973), 274.
43. J. Wayne Wrightstone et al., "Accountability in Education and Measurement Problems," *Test Service Notebook 33* (New York: Harcourt, Brace, Javanovich, Inc.), p. 5.
44. Stephen P. Hencley, "Impediments to Accountability," *Administrator's Notebook*, Vol. XX (Chicago: University of Chicago, December 1971).
45. M.M. Gubser, "Accountability As a Smoke Screen for Political Indoctrination in Arizona," *Phi Delta Kappan*, 55 (September 1973), 64-65.
46. Harold C. Hand, "National Assessment Viewed As the Camel's Nose," *Phi Delta Kappan*, 54 (April 1973), 539-41.
47. Ernest R. House, "The Price of Productivity: Who Pays?" *Today's Education*, 62 (September 1973), 65-69.
48. Jacob Landers, "Accountability and Progress by Nomenclature: Old Ideas in New Bottles," *Phi Delta Kappan*, 54 (April 1973), 539-41.
49. Allan C. Ornstein and Harriet Talmage, "The Rhetoric and Realities of Accountability," *Today's Education*, 62 (September 1973), 7080.

## CHAPTER 3

# ON THE NEED FOR CRITERION-REFERENCED RESEARCH AND DEMONSTRATION: A REACTION TO THE MODEL OF NATIONAL ASSESSMENT IN CITIZENSHIP

Joseph C. Grannis

Teachers College, Columbia University

In this paper I want to underscore first of all two related meanings of the term "model," one emphasizing the structure of what is represented by a model, and the other stressing the phenomenon of influence. Next, I would like to note the potential interplay between these two aspects of the National Assessment of Educational Progress (NAEP) model, with reference particularly to possible interpretations of the data in *Citizenship Reports 2* and *9*.[1] Last I shall suggest an alternative to National Assessment, taking note of the gains the project may have made, but at the same time orienting still more directly to political and education goals, in this case to the fostering of citizenship.

## The National Assessment Model as Representation and as Influence

This paper will dwell upon the obvious, and the distinction to be made here is no exception. Taylor's chapter (Chapter 2) describing the NAEP model and its uses itself points to both the representation and influence aspects of modeling, thus my starting point is one of clarifying what he has written.

When we speak of the "model" of National Assessment we refer partly to some representation of its basic components and their articulation. Exhibit 1 in Chapter 2 describes a succession of steps in the development, administration and dissemination of NAEP instruments and findings. The multiple connections drawn in the diagram suggest complications in this process, as the first-cycle sampling plan was modified in anticipation of scoring and analysis, exercise development looked forward to the utilization of information, etc., and as these steps might be repeated in successive cycles.

Exhibit 1 spells out, or differentiates, only steps of special concern to the assessor. A different diagram might elaborate the process between dissemination, utilization and the redevelopment of objectives, suggesting different channels of national, state and local dissemination; different uses of the assessment information as a function of different resources at these levels; and so on. Taylor's diagram reflects the putative neutrality of National Assessment vis-a-vis the uses of the assessment information. In various passages Taylor does suggest that this information calls for research, that it will be interpreted differently at state and local levels and that it will lead to interventions in different spheres — economic, curricular and so forth. Still, it is the basic stance of NAEP that these things be left to others to work out. From this standpoint Exhibit 1 represents the Assessment process accurately.

But a model is never neutral. Whether it has been drawn in physics or in politics and education, a model *models* a process of thought and action. In this simplistic way one can argue that *what National Assessment models is assessment*, and that complementary processes directed to changing what is assessed will be more an indirect, than a direct, result of National Assessment. It is not a trivial observation that the uses of the Assessment have in fact been, as described by Taylor, further assessment at the state and local levels. NAEP's powerful coordination of money and technical expertise with academic and political judgment partly accounts for this imitation. Equally, however, it is the specificity of the model with respect to certain processes, and its lack of specificity with regard to others, that accounts for the imitation. Indeed, the Assessment itself can be said to be an imitation of a process that has been most influential in American education in recent decades, the testing of individuals for purposes that are a great deal more diffuse than the testing operations *per se*.[2] The criterion referencing of the National Assessment exercises might be seen as a move beyond testing for its own sake, or beyond the parallel purposes of grading and selection associated with norm-referenced tests. Whether this move is realized, however, and the testing thus does not again become an end in itself, depends on other processes that the NAEP model does not explicate at present. If assessment is supposed to be linked more closely to teachers' accountability than it has been in the past, there remains the risk that accountability will be displaced upon the population tested.

Let me now repeat this argument at another level of what is modeled by National Assessment — its representation of what citizenship is and how one goes about observing and explaining its occurrence or nonoccurrence. Reasoning simplistically again, let us first suppose that acceptable performances on the citizenship exercises *are* the criteria of citizenship — i.e., that we would say someone *is* a good citizen if she/he performs in acceptable ways on the test, analogously to our saying that someone is a good driver if she/he passes a driving test.[3] How does NAEP allow us to explain the occurrence or nonoccurrence of these performances? Age, sex, parental education, color and location by region or size and type of community are the only variables that are systematically brought to bear on this question. One might surmise that age represents different degrees of exposure to citizenship education, but identifying whether it is explicit or implicit, or whether it stems from school, television or community, defies conjecture except on a most ad hoc, exercise-by-exercise basis. Sex might relate to role opportunity, parental education to kinds of discourse in the home, color to an individual's treatment by others and location to community needs and resources; but then we could scramble all of these conjectures too, given the lack of specification of what the variables link with. What we are led to, then, if only by default, is an emphasis on the population characteristics themselves as explaining acceptable and unacceptable citizenship performances.

Caplan and Nelson[4] recently reported that, of 69 data-based, psychological research studies of blacks that were abstracted in the first six months of the 1970 *Psychological Abstracts*, 82% were person-centered and only 18% situation- or environment-centered in their causal attributions. By far the largest category of variables Caplan and Nelson included under the heading "person-centered" was "group membership (e.g., black or white)," which included 48% of all the 69 studies. Person-centered categories, the authors observe, are more available and more easily exploited than situational or environmental categories. Discussing a variety of questions such as achievement motivation v. the structure of the economy as causes of unemployment, Caplan and Nelson concluded:

> Whether the social problem to be attacked is delinquency, mental health, drug abuse, unemployment, ghetto riots or whatever, the significance of the defining process is the same: the action (or inaction) taken will depend largely on whether causes are seen as residing within individuals or in the environment.[5]

Are we not dealing with a similar phenomenon? The effect of not specifying what age, sex, parental education, color and location mean experientially is our attribution of good and bad citizenship to these qualities *per se*.

The National Assessment citizenship reports do speculate at certain points about the meaning of the population variables. The following excerpts from *Citizenship Report 9* illustrate the tentativeness with which this is done:

> Thirteen-year-olds in the grade school group showed the greatest deficit of any group — 10% — on the racial-attitude exercises. Their responses to several questions describe where some of the deficits occurred, but don't explain why. . . . Perhaps as higher and higher proportions of the population get a high school education, the smaller the proportion who do not become more distinctive in certain ways (e.g., less accepting of other races).[6]
>
> Are black youngsters less willing to tell a nonblack interviewer what they believe about racial discrimination? Or are they really less aware of racial discrimination than other 13-year-olds? The balanced results on these questions show smaller deficits (by at least half) than do the observed results, as we discussed earlier in this chapter (see page 51). At least one of the other characteristics on which the results are balanced is thus associated in some way with the results reported for blacks. For example, perhaps the disproportionate number of blacks whose parents have little education hear less discussion about acts of racial discrimination in the world.[7]

These and a dozen similarly limited speculations aside, the main tendency in the reports is to leave the findings to explain themselves. Presumably research to be designed and conducted by other agencies will lead to the development of interventions that will in turn result in changes in the assessment results the next time around. What may defeat this, however, is the "psycho-logic" of the model. The structural emphasis of the model on population characteristics, rather than on the interactions of individuals with different environments, induces thinking in these same stereotyping terms by those who receive the assessment reports.

The language of the citizenship reports consistently reinforces stereotyping:

> The two upper levels of parental education, particularly the beyond high school group, excelled, as usual.[8]
>
> The typical performance of blacks at all four ages shows deficits of about 9% on all citizenship results combined.[9]
>
> The general picture is for performance on this goal to follow that on all citizenship results fairly closely. Thus the extreme rural and extreme inner city respondents showed the greatest deficit in relation to the nation as a whole, and the extreme affluent suburb respondents showed the greatest advantage. . . .[10]

Of course, this is all very ironic. The reports feature as many exceptions to the general trends as they can find, and the avoidance of discussing the significance of the general trends is presumably calculated not to risk offense. But the patterns in the test results that are associated with the population variables are clues to individuals' interactions with their environments. That our thinking cannot afford to rest with these clues is the thrust of the next section of this paper.

The last point to be made in this opening argument returns to the analogy between a citizenship test and a driving test. Simplistically again, we have to notice that National Assessment presents what individuals *say* in response to various paper and pencil, interview and discussion tasks as the basic model of what constitutes citizenship. Who can doubt that many more individuals know or will say that they should vote than actually vote in any election in this country — national, state or local? Or that more believe they should oppose discrimination in a park, and can say how to do so, than are likely to put this belief and knowledge into practice? Kohlberg[11] has emphasized Hartshorn and May's classic failure to find differences between delinquents' and nondelinquents' knowledge of "right" and "wrong" actions and has stressed instead the developmental level of the individual's justification of right and wrong. Kohlberg thus improves his capability of associating moral discourse with moral behavior, and, analogously, we might come closer in this way to associating citizenship discourse with citizenship behavior. The developmental approach, however, still leaves questions unanswered. Is it possible that those with "higher" levels of moral discourse are more capable of rationalizing their morality and immorality, and thereby more readily escape being branded delinquent or criminal? Our increasing awareness of "white collar" or corporation and political crime certainly gives credence to this question. Again, is it possible that different styles of language and discourse associated with socioeconomic or racial/cultural differences affect the response of an individual to a verbal representation of a moral, or let us say a civic, problem and affect equally an examiner's interpretation of the response? Here the questions we are raising with relation to what is known or suspected about moral judgment lead into the more general area of performance in a culturally standard verbal situation, in this case the formats that link testing and schooling so closely with each other. Labov's[12] finding that the task of talking to keep a rabbit from getting nervous elicited much more talk from children who spoke black English than questions addressed to the children by a sympathetic black interviewer illustrates this concern. Could all this have any bearing on who tends to "exhibit" less "knowledge and behavior considered desirable for citizens in our society"? Suppose the test was one of integrating a neighborhood, a bus or a lunch counter, or that it was challenging a school bureaucracy, or that it involved hiring or working or playing or generally living with someone who had been in jail, as opposed to saying how one would do any of these things?

Part of what is at stake here is the relevance of the knowledge required by the assessment to the particular contexts in which different individuals enact their citizenship. This question applies not only to, say, who is more likely to be oriented to the courts as an institution for settling disputes about money, but who is more likey to *have* to accept doctors and dentists of a different color from one's self. (Just to indicate that the question can cut both ways!) What I would emphasize here is that the Assessment is more likely to have content validity to the extent that it includes the enactment of citizenship in goal-related contexts. This then connects with the other part of what is at issue, the format or structure of the setting in which citizenship is to be observed, which on linguistic grounds alone, as I have tried to argue above, is more likely to be valid as it is oriented to a citizenship goal — i.e., to a citizenship goal other than acceptable performance on an exercise — not to mention how the Assessment's having acceptable and unacceptable responses for each and every exercise contributes to the cultural standardization of "advantages" and "deficits."

All of the above might seem to show that I have no use for informatioin about what individuals say, or that I think differences would vanish if enactment, instead of saying, was assessed. Actually, I would like to know about *both* enactment and saying, or the knowledge and feeling saying represents. The National Assessment exercises do in fact include self-reports of enactments, but, even accepting the reliability of self-reports, these are still no more revealing in and of themselves than are sheer statements of knowledge and feeling. How knowledge and feeling facilitate or inhibit enactment is a question that especially concerns educators, along with how situational conditions facilitate or inhibit enactment. Does the civics text knowledge represented by many of the exercises contribute to effective action in some political situations and ineffective action in others? Or, to rephrase a different question asked in *Citizenship Report 9*, what rules make most sense to different individuals participating in a common task?

Let us state the basic question still more generally. It is not who are the good citizens, but what are the conditions that contribute to good citizenship. Because NAEP is not designed to answer this question, however much it may speculate about it *ex post facto*, it does not model asking the question. We will now proceed further into the logic and "psychologic" of the problem this presents.

## The Meaning of Patterns in National Assessment Findings

National Assessment reports tend to discuss the explanations of specific anomolies associated with one or another exercise, but they avoid discussing the meaning of the larger patterns that hit the reader full in the face. Here we shall take note of some of the difficulties these larger patterns present. Our approach to this will be naive. We shall first orient ourselves to the data itself, asking what the patterns might be likely to suggest to the reader. Then we shall just begin to sort out some of the factors that are confused in these patterns, raising questions for an alternative program of research and demonstration.

### *Age as a Variable in Citizenship Performance*

In *Citizenship Report 2*, straight percentages of correct or acceptable responses at different age levels are shown for each exercise reported on. Thus the following results are exhibited:

| | 9 | 13 | 17 | Adult |
|---|---|---|---|---|
| Report that the police do not have the right to come inside one's house at any time they want and can give as a reason legal guarantees, or reasons concerning privacy and permission of occupant to enter(in own words).[13] | 20% | 68% | 90% | 83% |
| State that our legal system (courts, laws) is the means provided by government for settling an argument over money.[14] | | 50 | 70 | 87 |
| Last names of the persons now holding these offices. . . . : | | | | |
| President (Nixon) | 91 | 94 | 97 | 98 |
| Vice President (Agnew)[15] | | 60 | 75 | 87 |
| Could give at least 1 explanation of what fighting was about in country named | | 53 | 66 | 77 |
| 2 explanations | | 27 | 44 | 55 |
| 3 explanations | | 11 | 24 | 31 |
| 4 explanations | | 3 | 12 | 16 |
| 5 explanations[16] | | 1 | 4 | 7 |
| Opportunity to read a greater variety of viewpoints and information was stated as a reason why it might be good to have newspapers in a city written and printed by more than one company.[17] | 37 | 64 | 88 | 92 |

As these examples illustrate, the general tendency in the data is for the percentage of acceptable responses to increase with age, though with some reversals between age 17 and adult. This trend is alluded to in various specific contexts in *Citizenship Report 2*, for example in the following comment:

As seen above, adults consistently showed more knowledge about current conflicts than did the other ages. When asked in Exercise F4 to name some ways to avoid war, however, fewer adults than 13s and 17s named at least one way (77% of 13s, 88% of 17s and 65% of adults). Even 9-year-olds approached the adult achievement level, 60% giving at least one way to avoid war (Exercise F3). One explanation is that the adult view of the possibility of avoiding conflict is both more sophisticated and more "jaundiced"; a larger number of adults indicated that they felt war to be unavoidable and a larger number at the younger ages gave simplistic, but acceptable, answers ("stop fighting").

The number who gave at least three ways to avoid war demonstrates the more usual age trend of an increase in achievement up to age 17 with minor differences between 17s and adults (5% of 9s, 12% of 13s, 32% of 17s and 25% of adults). A similar 9-to-13 age trend is shown in Exercise F5, which supposes competition between the U.S. and Russia concerning territorial rights on Mars. Almost twice as many 13s as 9s (75% v. 45%) stated that the U.S. and Russia should discuss and settle these *matters before men* land on Mars.[18]

Notice that in the discussion above, as is the case elsewhere in the report, it is the departure from the "more usual age trend" that seems to call for explanation. How does one account for the main trend itself? Leaving aside for the present a consideration of how the exercises were constructed in relation to age expectations, and assuming the content or goal-referenced validity of the exercises, the conclusion that best fits the pattern in the results is that they reflect maturation and/or the general accumulation of knowledge and experience with age. This is obvious, is it not?

What is less obvious is the significance of the conclusion for citizenship education. Many, but probably not all, readers of this report will be aware that social studies educators in the 1950s (I choose this date for convenience of discussion) had occasion to assess the significance of similar findings arising out of a number of surveys of children's social concepts conducted during the Progressive Education era: Meltzer, 1925; Lacey, 1932; Pressey, 1934; Eskridge, 1939; Ordan, 1945; and Bates, 1947 — to name those I am most familiar with.[19] All of these studies included elementary school pupils at different grades, and several of them included secondary school pupils as well. All tested children's attainment of the correct or conventional meanings of social concepts at different ages, and all found progressions in correct attainment with increasing age. Interestingly, only elementary social studies educators, with the exception of Wesley and Wronski[20] (Wesley himself was author of a "Test of Social Terms" in 1932[21]), paid attention to these findings, directly or indirectly, to judge from a comparison of elementary and secondary social studies methods texts of the 1950s. At least in the decade before the impact of Sputnik, Bruner and the structure movement in curriculum, it was elementary school educators who were more concerned with the development of knowledge *over time* or age, while secondary school educators focused on its acquisition *at a given time* or age.

But which emphasis was correct? The elementary school educators stressed a progression of learning from concrete to abstract, the necessity of building up manifold experiences with the referents of a concept, and the importance of teachers' avoiding empty verbalisms and premature formalization. What precipitated out of this, however, was such an emphasis on accommodating to the presumed "natural" pace of concept attainment reflected in the progressions of attainment with age, that Bruner's declaration than anything could be taught in some honest way at any age, hedged though it was in Piagetian conditions, came as a shock, particularly to elementary school educators.

None of the social-concept studies cited above, nor any others conducted in the years before 1950, tried to assess children's capability of learning a concept at a given age by systematically attempting to teach the concept to children. Lacey did interpret an observed acceleration of concept learning at grade 3 as being the result of a more deliberately planned social curriculum than in grades 1 and 2. Though I do not know them, there may well have been many studies carried out during these years that weighed the relative merits of one method of teaching social concepts versus another, as a variety of more recent studies have done.[22] Who, however, has set out to teach social concepts to a criterion at a given age, i.e., to teach until the pupils reached the criterion, and in this way studied the methods that *resulted in* different children's learning *at that age*? We may often teach with reference to a criterion; but in the regular or the experimental social studies or civics classroom we typically abandon a unit of instruction to move on to another unit or to discontinue an experiment, while some pupils, *at least*, still have not met what we might hold as even a minimum criterion. Of course, this raises all kinds of other questions, about the desirability of convergent versus divergent learnings, etc., but that is beside the point here. What I sense is that there are a variety of factors, perhaps especially in social studies and civics, that contribute to a criterion-referenced system devolving into a norm-referenced system. One of these factors is the very knowledge that individuals' differences with respect to attainment of a criterion held to be "reasonable" for a given age will tend to diminish as more of these individuals attain the criterion past that age.

Without understanding the issue completely, I think there is an ambiguity in National Assessment's own position on this matter. On the one hand, we have been told that "outstanding local teachers familiar with each target age group (ages 9, 13, 17, adult) worked for weeks with our staff to break down each general objective in the most germane behaviors appropriate as goals for a given age group."[23] On the other hand, we learn that one criterion for exercise development was that "some exercises cover important attainments which nearly everyone is successfully achieving, some which very few people are achieving and some which a middling number achieve."[24] Does the latter aim, together with the finding that Assessment performances did in fact break down this way for each major citizenship goal,[25] imply that such a distribution should continue to be obtained in future assessments? Perhaps the aim of NAEP is to provide information that will encourage state and local striving toward full attainment of the goals for all individuals. Taylor, however, has observed a tendency for Assessment data to be treated as a national norm against which both state and local test performances are being compared (see Chapter 2). Given the statistical patterns within and between the age levels that National Assessment *models*, I think this is inevitable.[26]

National Assessment findings are likely to be taken as developmental norms unless several steps to counter this are taken. (1) The exercises would have to be broken down into their more truly developmental and nondevelopmental components; (2) the assignment of exercises to age levels would need to be done from the standpoint of what it is theoretically possible to expect any educable individual of a given level to achieve; and (3) experiments would have to be conducted to demonstrate that 100% of the educable individuals at an age level *could* achieve acceptable performances for the exercises at that level

This is a rather drastic prescription, with many problems inherent in it. Its function in this argument is to underscore the fact, recognized by NAEP but likely to be overshadowed by the pseudodevelopmental patterns in its findings, that National Assessment so far has not tried to take cognizance of what has in fact been deliberately taught, specifically in the schools, thus its findings should not be construed as suggesting limits to what individuals could be taught, or could learn, at the different age levels of the Assessment.

But we did admit that this has its problems. One of these is distinguishing what we called above more "truly" developmental from nondevelopmental components of learning. It seems to me, for example, that any 9-year-old could learn that policemen do not have the right of unrestricted entry to a home. At the same time, 9-year-olds' reasons for such a belief would be less sophisticated than 13-year-olds' reasons, focusing at the first level on, perhaps, the more direct consequences to the persons involved, and at the second on more general properties of the social system like the rights of individuals to privacy and the effects of not regulating entry into homes on other spheres of the social system. Instruction might then concentrate on all 9-year-olds attaining at least the first developmental level of justification, and on all 13-year-olds attaining at least the second level. Of course, in addition to the questions already raised above about the significance of different developmental levels, further questions would now come into play about the efficacy of direct instruction toward developmental goals.[27] Still, this would represent a considerable shift in the validity of the developmental problem.

Another problem stems from the question of whether a criterion in citizenship is more validly held for individuals or for the society as a whole. Consider the case for the latter standard first. Almond and Verba[28] have suggested that a political system might get overheated if too many of its citizens participated actively in the process of governance, beyond voting, and thus that there might be optimal levels of less than full participation for the stability of even a democratic system of government. (Almond and Verba were, indeed, rationalizing the levels of participation they found in American society.) Apart from whether or not one agrees with the value of stability in the system, this does lead one to distinguish between citizenship goals for the nation or for a region as a whole and goals for all the citizens of the nation or a region. Maybe only a few are needed to, for example, come up with many ways that war could be avoided or that discrimination in a park could be stopped. When *Citizenship Report 2* referred to "goals that only a select few were achieving,"[29] I wondered if a part of what this implied was that only a few are needed to achieve these goals.

The trouble with this reasoning, however, is that the many who do not achieve the goals in question may include precisely those whose interests are most at stake, viz., those most likely to be frontline soldiers in a war or most likely to be excluded from a park. Once we allow that less than all need to achieve one goal or another for the health of the body politic as a whole, we run the risk of playing into just those features of the system that presumably account for some groups in the population being consistently "disadvantaged" no matter what the goal in question is. Furthermore, many of the goals set by National Assessment pertain to the rights and obligations of individuals in their everyday relationships to one another and to the law. Anyone of a relevant age who does not know that the courts are available for the resolution of a dispute over money might be disadvantaged in the event of such a dispute. Similarly, anyone who discriminates against another on the basis of his or her group identification contributes to the denial of the rights of the other to be treated as an individual.

A third problem arises from the fact that, especially if, in line with the position advocated earlier, instruction with respect to a particular goal at a given age was pursued until *all* individuals so instructed reached the goal criterion, there might not be time enough to reach all the goals held to be desirable for individuals at that age.

Still another problem is that it cannot be assumed that school is the most appropriate setting in which to intervene to try to bring individuals to a criterion. But then there is no way of knowing from the present design of NAEP what part the schools have played in the performances on almost any of the assessment exercises. Even specifying which criteria the schools could effect would be progress from this standpoint. Especially for those criteria that we have characterized as "enactments," it *might be that* school would have to be coordinated with other agencies or institutions, family, local government, citizens groups, television and so on. National Assessment may assume that efforts of this sort will be an outgrowth of the publication of assessment results. It would be a more likely outcome if the contribution of such efforts to the attainment of goals was specified.

What all of this reasoning drives me to, then, is a different kind of neutrality from that which is modeled by NAEP. Instead of publishing results that encourage normative thinking in the face of virtually no knowledge of the conditions of individual citizenship learning and development, an alternative strategy should concentrate on specifying the conditions that make full attainment possible for one or another criterion at a given age level. The freedom of state and local agencies or institutions to emphasize those goals that mattered most to them would thus be an enabling one, rather than the spurious freedom that sheer ignorance affords us.

### *Sex, Parental Education, Color and Location by Size and Type of Community (STOC) as Variables in Citizenship Performance*

The logic with respect to these variables is essentially the same as it has been above, though I will be repeating it in somewhat different terms. First, let us attend again to the "psychologic" of the matter. The very familiarity of National Assessment findings, especially with respect to parental education, color and location, may tend to give them a normative significance, i.e., to imply that the obtained patterns are what we should continue to expect. For me, at least, this stereotyping is reinforced by the technique of comparing group performance levels with national levels, rather than with the criterion of 100%, so that there will regularly be groups with "advantages" and groups with "deficits," *relative to each other* (shades of norm referencing?), even when, say, it is 60% of one group and 50% of the other that has reached a criterion. Be that as it may, the familiarity of the results that I am referring to is their comparability to the findings of any number of assessments, from the Army Alpha intelligence testing program that might be said to have started it all, to the Coleman Report and still more recent studies. The sheer accumulation of these findings has contributed to their reification in recent years, so that in the Jencks analysis socioeconomic status and color appear to be almost intractable constraints on education, or at least on education by the schools.

Let me focus this thinking in terms of the Coleman and Jencks analyses, raising a few questions at a very superficial level. First, these studies base their (somewhat differing) conclusions about the limited capacities of schooling to affect achievement on achievement defined in very close relation to general cognitive skill, with all of the cultural loadings that the phrase "general cognitive skill" implies. As Jencks himself observes, the Coleman tests of Verbal Ability, Nonverbal Ability, Reading Comprehension, Mathematics Achievement and General Information intercorrelated so highly that "the student who did well on one test and poorly on another was quite exceptional."[10] Assuming for the purposes of this argument that socioeconomic and cultural factors reflected in the school will largely determine the school's effect on general cognitive skill, it remains possible that schooling can have a more independent effect in areas of achievement that are more specific to school instruction and, conversely, less generally diffused in the society at large. The International Education Assessment's (IEA) current research on factors affecting achievement suggests, for example, that the effectiveness of science instruction is substantially more independent of nonschool factors than is the effectiveness of instruction in reading or arithmetic.[11] Pursuing the principle that might explain this, we first have to account for IEA's finding that their Social Studies/Civics Education test results behave more like reading and arithmetic than like science!

The IEA researchers themselves attribute their findings to the permeation of citizenship knowledge, like the three R's, throughout life outside the schools. It makes sense from this standpoint that the citizenship results would reflect varying access to, and occasion or power to use, the knowledge that the tests incorporate. But is not science also diffused throughout life in a modern society? Of course this is the case. One must reason then that science education *and* the science test exercises have been couched at a higher or more specialized level of knowledge than is reflected in everyday life, and that citizenship education and/or the citizenship test exercises have not been defined at the same relatively high or specialized level. Studies of the learning of higher mathematics, as opposed to the common mathematics the Coleman (re-analyzed by Jencks) and IEA studies were concerned with, do show that differences in schooling make a difference. What is needed, then, is to demonstrate a similar effect in citizenship.

What might be meant by "higher" or "more specialized" citizenship knowledge? One need not think of it simply as, say, secondary school or college level information and concepts, but rather as knowledge that is couched at a higher or more complex level than seems to be reflected in citizenship achievement at any given age under conditions other than systematic instruction. Thus one would aim to teach at a given age what NAEP or some other survey shows is *not* "normally" achieved at that age, or perhaps not even at a subsequent age.

Another, complementary way of construing what might be meant by higher or more specialized citizenship knowledge stresses, instead of official or ideological doctrine, a more skeptical, probing interpretation of citizenship. Neither NAEP nor IEA has included *knowledge* exercises of this sort, although the IEA instruments do include attitude or opinion items that question the economic and political systems of the society. Can we say that the schools should also aim to teach to individuals of a given social background that which is not "normally" known to persons of that background in our society, be they rich or poor, and should cultivate questioning in that social sector of a sort not "normally" realized in that sector?

Arnoff's[12] investigation of factors related to the ability of children in second, third and fourth grade to comprehend concepts of government points in the first direction suggested above. Arnoff designed a five-week government curriculum to include (though not exclusively) concepts not ordinarily included in instruction for these grades — i.e., not included in social studies textbooks at these grade levels. Arnoff's results clearly showed the effect of instruction. Seventy-five percent or more of second graders, for example, learned 23 new concepts of local, state and national government: property tax, split ticket, subpoena, judge, etc., most of which were not included in second-grade social studies textbooks. Furthermore, social class tended not to correlate with more or less learning of new concepts in this experiment, though mental age as defined by an intelligence test did correlate with new learning.

A field trial of the American Political Behavior (APB) course developed by the High School Curriculum Center in Government at Indiana University[13] points partly in the second direction indicated above. The APB course aimed to teach "facts and ideas about politics that have not been part of typical social studies curricula,"[14] for example, facts about the proportions of different income groups that vote in American elections. Differences between experimental and control groups' pretest/posttest knowledge gains were clearly demonstrated in all nine communities involved in the field trial. On the other hand, comparable differences in political science skills achievement were demonstrated in only four of the nine communities, and effects on students' attitudes were demonstrated in none of these communities. Furthermore, while there was some variation of student background characteristics, the communities involved did not include the rural and inner city extremes identified in National Assessment, and almost all of the students were white.

One should not suppose that systematic instruction could eliminate the effects of population or student background characteristics, or indeed that this adequately represents what is desirable. The effects that National Assessment makes us principally aware of are those that stem from restrictions on different groups' access to and control of information; if citizenship education should aim to minimize these effects, still this would surely entail interventions outside the schools, and research and demonstration would need to be directed this way as well. At the same time, the schools must recognize the identities and priorities of different social groups; paradoxically, whatever equalization the schools accomplished with respect to the distribution of information and skills might contribute in some ways to heightened affective differences between groups.

A second problem in the Coleman study, one that the Jencks analysis only partially rectifies and that National Assessment simply ignores, is the question of comparability of school environments, in terms other than socioeconomic status (SES), between and within different locations. It is well known that Coleman compared whole systems as to library size per-pupil expenditure, level of teachers' formal education and so forth, whereas these conditions might well vary between schools within a system. Jencks used Project Talent data to make these comparisons between individual high schools, and still obtained no effect independent of nonschool factors. The IEA science findings *do* show such an effect, not for all the variables that might be thought to be relevant, but for "the opportunity to learn, the student time in hours per week and cumulative years, the curriculum emphasis and the additional years of postsecondary preparation of teachers."[35] In none of these analyses has classroom climate or methodology seemed to account for differences in students' achievement.

My own position is that one must compare not just programs or schools or even classrooms in order to specify the environmental conditions that affect schooling's contribution to achievement. Rather, it is necessary to examine different settings *within* classrooms, or within *school-related* locations outside the classroom, in order to make headway with the question of educational environments.[36] In my approach to this problem over the past few years, I have found it useful to distinguish between high teacher control, joint teacher and learner control and high learner control of each of various conditions within one or another classroom setting (or subsetting) and thus to be able to ask how much time the students in a given classroom spend under different conditions of control. Task options, pacing, teacher-learner and learner-learner interaction, teacher adaptiveness and task-performance criteria are some of the conditions of settings that can be observed to vary with respect to control. So-called "open classrooms," for example, can be seen to vary among themselves in the proportions of time that students spend under these different conditions of control, whether one is comparing time for aggregates of students between two or more classrooms or is comparing time for individual pupils within a single classroom.

This approach further distinguishes between different types of education purposes: the transmission of knowledge, tradition and experience to all learners in a setting alike; the operationalization of partly common, but also partly individuated, competencies or skills and concepts; and the cultivation of individual and collaborative venture, or exploration, problem solving and expressive composition or construction. Intellectual, social, physical and emotional aspects of learning and development are equally involved in each of these categories of purposes.

I have hypothesized both that the effectiveness of an educational setting will be partially dependent on the internal consistency or congruency of its controls, and that each broad type of education purpose will be realized best in a different type of setting, viz., transmission in a high teacher-control setting, operationalization in a joint teacher- and learner-control setting and venture in a high-learner control setting. It is apparent, then, that I believe that what has variously been called "classroom climate" or "classroom method" will someday be understood to relate to achievement, or to type of achievement, even though the results of research on this to date have been very confusing.

This report has emphasized education in the schools, but it has recognized at more than one juncture that going beyond learner or citizen population variables to the interactions of individuals with their environments would entail research and demonstration outside of the school as well. What are the effects on individuals' acknowledgment of racial discrimination, of this acknowledgment's actually contributing or not to ameliorating the discrimination? How is one's orientation to a legal system for settling an argument over money affected by one's having, or anticipating having, money enough to invoke the legal system in a dispute? Does involvement in the affairs of a responsive public agency or institution lead to greater participation in other public or community affairs?

My paradigm for these questions, and many more, is Kohn's[37] research on the conditions that determine fathers' values for their children. Social class, defined as occupational position and education, does relate in Kohn's data to whether parents tend to value self-direction or conformity to external authority — higher social class being associated more with the first and lower social class more with the second of these value clusters. When the analysis controls for the degree of self-direction or autonomy that the fathers experience at work, however, social class differences tend to vanish, lower-class fathers who experience autonomy in their jobs valuing autonomy for their children as much as middle-class fathers who experience this autonomy do. In other words, while social class is an approximation to the conditions that determine fathers' values for their children, the experience of fathers in their work describes these conditions more exactly. Is it not possible that research could similarly track down conditions associated with, but still independent of, sex, parental education, color and location that would more exactly explain the associations of citizenship achievement with these variables? Further research could then be directed toward changing these conditions experimentally, so that, just as we might demonstrate in what way achievement other than that which is "normally" associated with different age levels is possible, likewise we could show how achievement other than that "normally" associated with different sex, parental education, color and location statuses is possible.

## An Alternative to the National Assessment Model

From various directions, our argument converges on the desirability of emphasizing person-environment interactions and demonstrating the achievement that is possible under varying person-environment conditions. Consider one further vantage point that differs somewhat from ours so far. Etzioni,[38] in a discussion of organizational analysis, criticizes that assessment of organizations that focuses on goal attainment. "One of the major shortcomings of the goal model is that it frequently makes the studies' findings stereotyped as well as dependent on the model's assumptions."[39] An organization frequently does not reach its goals effectively and often has goals other than the ones it claims to have. It would be more useful, Etzioni concludes, to treat goals as cultural entities in themselves and to ask how various internal and external conditions contribute to the realization of observed goals.

I do think NAEP's attempt to define citizenship goals and to describe performances that represent the achievement of these goals is valuable. The fact that National Assessment anticipates modifying the goals and exercises to reflect changes in priorities between assessment administrations indicates at least some awareness of the goals as "cultural entities." But what is served by publishing only those statements of goals that represent the "consensus" of those consulted? Would not the publication of disagreements over the definition of goals, particularly disagreements stemming from different subcultural interpretations of citizenship, add further to the demythification and destereotyping of citizenship goals?

What stands in the way of this process is, once again, the Assessment model, its structure and assumptions. Let me claim intuitively that a national test of citizenship simply cannot be consistent with cultural and political pluralism. National Assessment *is* first and foremost a test and as such depends upon conformity to its goals for its authority.

An alternative to National Assessment would conduct research and demonstration in specific relation to citizenship goals but would recognize that these goals have different meanings and priorities for different individuals and groups. The alternative should, indeed, explore these differences explicitly, as an understanding of them would be essential to any application of the research and demonstration.

What I have in mind first of all, then, is a pluralism of inquiries to replace the monolithic National Assessment, inquiries that could be drawn upon differently by different agencies or groups or rather, since anything can be drawn upon in this way up to a point, that *encouraged* different agencies or groups to order their efforts according to what is most meaningful and important to them.

An alternative to National Assessment should be truly criterion oriented. Its aim should be to use analysis and intervention to bring the observation of performances as close as possible to criterion, *allowing* that different agencies and groups will draw differently on this research and demonstration in subsequent applications of it.

Many separate studies would be involved, but their intent would be the same. The basic paradigm would be that which shows convergence on the achievement of goals as the result of reducing differences between groups, or aggregates, of different statuses, by specifying the conditions associated with these statuses that originally account for the differences in goal achievement. Is the seemingly lesser awareness of racial discrimination on the part of blacks a result of unwillingness to disclose this awareness in a test or interview? Then perhaps the difference would diminish as observations were conducted in settings in which the respondents felt more in control. Are parents of lesser educational attainment less involved in the politics of their children's schools because of a feeling that school had not been responsive to them as children? Comparing groups on the basis of this feeling might reduce the difference attributable to status and intervention through the schools to respond to alienated parents might increase the parents' involvement in their children's schools.

Probably hundreds of studies that throw light on the specific conditions contributing to citizenship have already been reported. These could be indexed by goals, and again by conditions, so as to make this knowledge available to policymakers at different levels. Somewhat in example of this is a document prepared by the Social Research Group at The George Washington University, *Research Problems and Issues in the Area of Socialization*, 1972, part of which analyzes what is known about the development, the determinants and the changing of intergroup and intragroup attitudes and behaviors. Of course, such an analysis points as well to what is *not* known. The following quotation from the Social Research Group's report represents a juncture that is frequently arrived at in their analysis:

> The significance of some of these findings on cooperation is far from clear. We still do not know how cooperation determines intergroup attitudes and behaviors or its role as a factor in achieving a successful ethnic and social class mix. What does seem clear from the research is that cooperation sometimes leads to better intergroup relations, although much work is needed to determine the conditions under which cooperation produces an enhancing effect.[10]

Much work does need to be done. Some of it consists of formal studies in which the investigator attempts to control the principal variables and to predict outcomes precisely in advance. Often, however, these studies are too rigid to be able to deal with the unintended factors that enter into every complex action — which may partly account for formal studies failing to confirm their hypotheses. Action research, in which the action is typically guided by more evolutionary goals and in which "real time feedback" continuously regulates the participants' activity in relation to goals, compensates in flexibility for what it may lose in control. A formidable agenda of both formal and action studies, then, must be undertaken, all of it referenced with respect to goals and conditions alike.

Perhaps the case could be made that developing a knowledge base for citizenship education in this way would still require a determination of national performance levels as in NAEP. In my opinion, however, this would be distracting at the least, and possibly destructive. An alternative way of construing the "national" significance of a study would be that it deals with problems that are best approached with national resources, while a more local study deals with problems for which local resources are adequate, or, to put it differently, for which *only* local resources are likely to be appropriate. Thus a study could have national significance even if it did not have a national sample (which is not the same as saying it would not be carried out in a variety of communities). The effects of income maintenance on citizenship performance might be construed as a national question because only the federal government has the money to maintain incomes. The effects on citizenship performance of collaboration between schools and other community institutions or agencies might be construed as a local question because only local individuals or groups have the power to bring about this collaboration.

Clearly a citizenship research and demonstration program could go on forever, insofar as our awareness of what we do not know multiplies at the same rate, at least, as does our knowledge. Then National Assessment equally could continue forever, since the goals it sets will probably change in like manner. Substantial human and material resources have been invested in NAEP but this is not the only, or even the chief reason, to think that the project *will* be continued. The technology of the Assessment, as of testing more generally, peculiarly lends itself to the collusion of academic ingenuity and political decision making. Were the questions this paper raises foreseen? Probably some of them were, and others were not. If the present form of National Assessment is itself the result of compromises meant to render the project less dangerous politically, this does not bode too well for the alternative that I propose. On the other hand, what are the political consequences of what National Assessment hath wrought? These too will have to enter into the equation.

## *NOTES*

1. See Appendix D, References 5 and 3.

2. Jan Duker and Mary A. White, "Models of Schooling and Models of Evaluation," *Teachers College Record*, 74 (February 1973), 293-307, makes the connection between testing and selection a relatively diffuse, if still consequential, process.

3. This analogy is drawn, with much the same intent, in Naomi White, *Statewide Testing Legislation for Minimal Skills: An Exploration of Some Issues* (Educational Policy Research Center, Syracuse University, 1973).

4. Nathan Caplan and Stephen D. Nelson, "On Being Useful," *American Psychologist*, March 1973, pp. 199-211.

5. Ibid., p. 201.

6. See Appendix D, Reference 3, p. 25.

7. See Appendix D, Reference 3, p. 57.

8. See Appendix D, Reference 3; p. 37.

9. See Appendix D, Reference 3, p. 46.

10. See Appendix D, Reference 3, p. 92.

11. Lawrence Kohlberg, "The Development of Children's Orientation Toward a Moral Order," *Vita Humana*, 6 (1963), 11-33.

12. William Labov, "Academic Intelligence and Black Intelligence," *Atlantic Monthly*, 229, No. 6 (June 1972), 59-67.

13. See Appendix D, Reference 5, p. 30.

14. See Appendix D, Reference 5, p. 41.

15. See Appendix D, Reference 5, p. 54.

16. See Appendix D, Reference 5, p. 87.

17. See Appendix D, Reference 5, p. 104.

18. See Appendix D, Reference 5, p. 84.

19. H. Meltzer, "Children's Social Concepts: A Study of Their Nature and Development," *Teachers College Contributions to Education*, No. 192 (1925); Joy M. Lacey, "Social Studies Concepts of Children in the First Three Grades," *Teachers College Contributions to Education*, No. 548 (1932); L.C. Pressey, "A Study in the Learning of the Fundamental Special Vocabulary of History from the Fourth Through the Twelfth Grades," in *Tests and Measurements in the Social Sciences*, eds. Truman L. Kelley and A.C. Krey (New York: Scribner's, 1934), pp. 155-218; T.J. Eskridge Jr., "Growth in Understanding of Geographic Terms in Grades IV to VII," *Duke University Research Studies in Education*, No. 4 (1939); H. Ordan, *Social Concepts and the Child Mind* (New York: King's Crown, 1945); F.L. Bates, "Factors Related to Children's Understanding of Social Concepts," (Doctor's thesis, University of California, 1947).

20. Edgar B. Wesley and Stanley P. Wronski, *Teaching Social Studies in High School* (Boston: Heath & Co., 1958).

21. Edgar B. Wesley, "The Wesley Tests in Social Terms," in *Tests and Measurements in the Social Sciences*, eds. Truman L. Kelley and A.C. Krey (New York: Scribner's, 1934), pp. 219-226.

22. For references to these, see the annual reviews of research published in *Social Education* since 1966.

23. Vincent N. Campbell and Daryl G. Nichols, "National Assessment of Citizenship Education," *Social Education*, *32* (March 1968), 280.

24. See Appendix D, Reference 5, p. 5.

25. See Appendix D, Reference 5, p. 6.

26. Ebel has made the interesting observation that the percentage system of grading that prevailed 40 years ago and before was a criterion-referenced system. He discusses some of the reasons for its yielding to a norm-referenced system in Robert L. Ebel, "Criterion Referenced Measurements: Limitations," *School Review*, 79 (February 1971), 282-288.

27. Lawrence Kohlberg and Rochelle Mayer, "Development as the Aim of Education," *Harvard Educational Review*, *12* (November 1972), 449-496.

28. Gabriel A. Almond and Sidney Verba, *The Civic Culture* (Boston: Little, Brown and Company, 1965).

29. See Appendix D, Reference 5, p. 5.

30. Christopher Jencks et al., *Inequality* (New York: Basic Books, 1972), p. 55.

31. "Scholars Must Weigh Education Data from Twenty Countries," *Report on Education Research*, 5 (December 1973), 3-4.

32. Melvin Arnoff, "An Investigation of Factors Related to the Ability of Children in Grades Two, Three, and Four to Comprehend Concepts of Government" (Doctoral dissertation, Ohio University, 1963).

33. John L. Patrick, "The Impact of an Experimental Course, 'American Political Behavior,' on the Knowledge, Skills, and Attitudes of Secondary School Students," *Social Education*, 36 (February 1972), 168-179.

34. Ibid., p. 168.

35. "Scholars Must Weigh Education Data from Twenty Countries," *Report on Education Research*, 5 (December 1973), 3-4.

36. Joseph C. Grannis, *Columbia Classroom Environments Project Final Report*, Contract No. OEC-0-71-0593 (Teachers College, Columbia University, February 1973).

37. Melvin L. Kohn, *Class and Conformity* (Homewood, Ill.: Dorsey Press, Inc., 1969).

38. Amatai Etzioni, "Two Approaches to Organizational Analysis: A Critique and a Suggestion," *Administration Science Quarterly*, 5 (1960), 257-278.

39. Ibid., p. 258.

40. Joyce B. Lazar and Barbara J. Sowder, *Research Problems and Issues in the Area of Socialization* (Washington, D.C.: The George Washington University Social Research Group, 1972), p. 82.

# CHAPTER 4

# EVALUATING SOCIAL STUDIES AND CITIZENSHIP EDUCATION: SOME ALTERNATE APPROACHES

Michael Scriven

University of California, Berkeley

There is tremendous overlap between the areas of social studies and citizenship education — well-illustrated by the overlap in the goals and objectives developed by the two different institutions, the Educational Testing Service (ETS) and the American Institutes for Research (AIR), that contracted with the National Assessment of Educational Progress (NAEP) to develop specific statements. I shall not make any great effort to separate them here because our concern is with evaluation models and only secondarily with specific content — and very similar problems arise in both areas.

Let's call the NAEP approach, well-described by Bob Taylor, the first approach to evaluating social studies education. I shall describe three other approaches very briefly and suggest a synthesis. Then I shall look in somewhat more detail at certain features of the National Assessment approach and Bob Taylor's comments. I'll begin by focusing on the area of citizenship values, understanding of law and due process, etc., because it's more important than geography and in worse shape.

## Alternatives to National Assessment

A second possibility would be a comparative approach in which direct international comparisons, for example, were sought. This process would require a substantial but not complete revision of the item pool in order to make really direct comparisons; but some direct and many indirect ones can be made using the data we do have from the Institute of International Education studies and others. These comparisons are extremely important — even more than they were in the math studies, for example — in showing what can be done as well as what is being done. The citizenship-/social studies areas are not so abstruse that one can reasonably suppose them (or most of them) to be beyond the grasp of a substantial majority of pupils. Of course, there are important differences between countries that would make it necessary to proceed with caution, in inferring from what has been done elsewhere to what could be done here; but the differences are not such as to make attempts (experiments) absurd which is all we could justify to start with anyway. Of course, too, tremendous chauvinism is associated with the citizenship/social studies area, and there would be many who would condemn any attempt to match the performance of other countries *per se*. However, the merits of that argument seem as slight here as in, for example, the automobile or the psychiatric or the adult education field. With General Motors switching to the Wankel engine, acknowledgedly copying Mercedes in styling and suspension and Ferrari in design (with the 1975 models), one can hardlly argue that the experts can't see any transferability of foreign ideas. Similar cases are well-known in the other fields mentioned. If we are interested in reducing the level of antisocial activities and basic ignorance about constitutional and other rights, and about human nature — which is what citizenship/social studies is all about — then it seems appropriate to look for possible improvements wherever we can find them. There is plenty of evidence in the comparative education studies to date to suggest that we could do better, but we do need more precise comparisons. Telling us whether we're doing as well as we could is one of the functions of evaluation, and it isn't done very well by NAEP.

Another type of comparison that would really be significant would involve comparisons between the performance of pupils — still measured mostly on paper and pencil tests, as in National Assessment — attending schools with radically different approaches to citizenship/social studies. The contrast between those with a conventional curriculum and those using some of the alternative approaches — e.g., Project Social Studies materials — might be illuminating, and the discovery that there wasn't a contrast would also be illuminating. Of course, the Assessment does make some comparisons, e.g., between performance of black and white pupils. It's a little hard to alter people's color; it's less hard to introduce new curricula or methods. One might put it this way: NAEP and most state assessment programs are pretty good photographers, but not very good buying guides. For that you need the relevant comparisons, viz., those between the available options.

A third possibility involves switching to a very different kind of item, albeit still a paper-and-pencil (or vocal) test. Here we'd go to something like the Social Issues Analysis Test, where the item might present a page-long newspaper editorial or a dialogue and a series of rather searching questions on it, of a wide range of difficulty. We might allow a lot of time for this — perhaps an hour. An essential selection criterion for items would be novelty: typically they would be hypothetical cases to which the "approved" answer has not yet been identified by adults with whom the pupils interact (to rule out parroting). The reasons for shifting to this kind of item are (1) that the few items like this that have been released reveal the most appalling incompetence in operating with the simplest constitutional or moral principles (e.g., freedom of speech), and it is now most urgent to clarify the real situation; (2) although the analysis of responses would require formidable training and talent on the part of the scorers (since much reading between the lines would be necessary) there would be a corresponding increase in the significance of the results. Instead of telling us where the pupils are at, this kind of test can tell us where they are capable of moving to, which is our only hope (given the abysmal level of performance at the moment). The page of dialogue can involve argumentation and can call for evidence of understanding the steps in the argument and their effect on the reader (e.g., by using interspersed questions and indelible markers, etc.). Despite the use of hypothetical situations, the responses are much more likely to be realistic here than in the present items, where the use of stereotypes by the student can easily provide a facade of answers that tell us nothing about the probable response to a new case. In short, bad though the present answers are, they may well give huge overestimates of the merit of the respondents, which leads us back to reason (1) above.

A somewhat radical extension of this approach leads to a fourth evaluation, which would move into the field and away from pencil-and-paper tests and use the best skills of the anthropologist and the sociologist (besides those of an extremely acute content analyst) to identify the values of various age levels and adults in our society from a study of their communications and decision processes. To take an extreme, but extremely important, example — during the hearings before the Ervin committee we were presented with a very detailed picture of the level of moral analysis and citizenship behavior of the White House staff. The addition of the tapes has made this a very complete data source for the kind of question I'm raising here. Similar analyses can be done of the discussion at the school board and in the local press of a proposed decriminalization order to a city police department (and of subsequent events), or of the discussions in an eight-grade classroom of a proposal to vest disciplinary powers in the students (and of the subsequent events). These analyses are tricky; there are few analysts presently equipped to deal with them objectively — but oh, what a treasure trove for the evaluator is there! Here we can bypass the problem of test invalidity; here we are dealing with real actions and perceptions. Despite the massive media coverage of Watergate, I never saw any analysis of the significance of the conceptions revealed by Haldeman and Ehrlichman on the stand. Most people got the feeling that they were "sort of morally blind," that they were abusing their power. But consider Ehrlichman's justification of the burglary of Dr. Fielding's office: "As we saw it, it was as if you learned that a map was stored in the vault of a D.C. bank that showed the location of a bomb that would blow up the whole of the district the next day; wouldn't *you* think it was justifiable to 'break and enter'?" [paraphrase]. There were a dozen similar examples. I think there is more information about the evaluation of citizenship/social studies in U.S. schools and homes in those passages than in all the corresponding test results from National Assessment. This was not just one aberrant lawyer speaking. This was a line of argument — grotesquely irrelevant though it was — that readily persuaded almost everyone on whom it was tried by someone coming from the White House. We did not learn about Watergate from someone who had a better education in citizenship/social studies than Ehrlichman; we learned about it from a black nightwatchman doing his job well — for which he was essentially blacklisted.

There is a recurrent tone in the Watergate discussions at every level — media, Congress and neighborhood — and the same note can be detected in the discussions of any other widely discussed moral issue of our time, such as drug law and enforcement, "excess" profits by oil companies, etc. That tone is naiveté — and from our point of view, particularly naiveté about the psychological nature of mankind, society and morality. We need more careful evaluation than we have yet had to determine whether this impressionistic reaction is ill-based or not. The third and fourth methods described here use simple enough procedures, which we have often applied in evaluating competency in other areas of interest, e.g., in testing cognitive, mechanical and administrative skills. I believe they deserve more serious application in citizenship/social studies, where we have so far — with regard to our own society — alternated between oversimplified paper-and-pencil tests and overemotional social documentary.

One feature of the field-study or anthropological approach, which deserves some stress, is that it does not begin (or does not need to begin) with the massive effort involved in developing goals. There is something slightly inappropriate about that effort for an evaluation task, it seems to me; it is exactly the right activity for developing a new curriculum, but that is hardly what NAEP was supposed to be up to. (It's perhaps not too surprising that considerable opposition to National Assessment arose from those who felt that it was attempting to impose a monolithic citizenship/social studies curriculum on U.S. schools; the complaint might seem stupid at first glance but on second thought reflects some sensitivity to a significantly possible outcome, school politics being what they are.) It seems plausible enough to argue that you can't set up tests until you know what they are *tests* for, and what they are tests *for* — i.e., the goals of citizenship/social studies education. But that's an error, as we'll see in the next section. Here I'll just stress the existence of an alternative approach. One could have had a team analyzing adult behavior in the citizenship area for deficiencies by identifying the optimal feasible behavior in the situation in question and extracting the discrepancies. After a long search, one would then classify the discrepancies and set up the assessment program to determine the extent of these deficiencies in the population. This involves no reference to the goals of citizenship/social studies education, though such could be inferred from it; it short-circuits that concept.

So the three models I am proposing might be called the *comparative*, the *simulation* and the *anthropological* models. They are mere sketches here, of course, but I believe they do serve to open our minds to the existence of rather different approaches to evaluation of citizenship/social studies education — possibly they will serve as useful targets for discussion. We have become somewhat fixated on the "standard" model of assessment, and we have invested in it very heavily (see annual reports of Educational Testing Service

[ETS] Center for Statewide Assessment). I think we have become too rigid in using this model, and I see no reason why some diversion of resources could not be made to include at least some of the other models I have described.

But that's not the only possible way to change. There are major changes in the NAEP model that deserve consideration and that could also produce an "alternate form," which could be used alongside the continued use of the present forms (desirable for obvious reasons). We'll turn to these in the next section.

## Changes in National Assessment

A tremendous price was paid for "political" acceptability of the NAEP approach, and this may well have been the right decision. However, there is some point in talking about ideal ways of evaluating, and even the feasibility question is probably due for reconsideration. The two big trade offs (or sellouts, depending on how radical one feels this morning) are:

1. Restriction of goals to those "accepted as an educational task by the school,"[1] or "acceptable to most educators and considered desirable teaching goals in most schools."[2] A further restriction was to goals that were "considered desirable by thoughtful lay citizens."[3]

2. Restriction of items to those that most states liked. In some cases, it is clear, particular states would not accept certain items in the version of the test forms used within their boundaries, and to avoid becoming widespread, compromises had to be made on other items. Of course, citizenship/social studies were the areas hit hardest by this constraint.

These are serious limitations indeed. If schools and states can vote on the standards by which they are to be judged, we are simply going to lose some very fundamental criticisms. The process actually gave the veto power to each of three groups — scholars, educators and lay people. That's a pretty tough obstacle race for an objective to get through and some pretty crucial ones didn't make it, especially — unfortunately — those that would most acutely test the moral sensitivity of students on controversial issues.

These restrictions might be relaxed after new consultations, or they might be bypassed using the field-study approach described earlier. One way or another, they are barriers to a full evaluation.

One does not judge the education of lawyers or doctors by asking the law school or medical school for criteria (or letting them veto external lists). One judges it by a careful analysis of the performance of the professionals in the field, using the testimony of clients and coworkers who see that work — an analysis that looks not only at deficiencies, which will always be with us, but also at the question of whether these deficiencies are the kind that could have been removed by education, preferably an education that is fiscally and temporally realistic.

The schools are permeated by a number of unfortunate ideologies in the citizenship/social studies area, ideologies that are tremendously destructive to reasonable citizenship/social studies education, and completely fallacious. In the light of these ideologies, educators reject certain kinds of goals for citizenship/social studies; yet these include many goals of the greatest importance. A couple of examples may suffice to illustrate the point. The fact-value distinction, and the associated ideology of value-free science, is pervasive among educators (and many scholars and thoughtful lay people). Hence they will not accept goals that assert the objectivity and factuality of certain moral standards, and the falsity of others. Indeed they go further and require (i.e., did accept) objectives like SSIIA, 17-A: "Distinguish among definitional, value, and factual issues in a dispute."[4] This is, of course, the thin edge of the relativistic wedge. If one can't say that it's a *fact* that Ehrlichman *improperly* approached Judge Byrne, then ethics is indeed a travesty; but of course it *is* a value claim and if these are exclusive categories, it can't be both at once.

There are other glaring omissions in the objective lists concerning the foundations of ethics, the relation of ethics to religion, to conscience, to the law, to custom and convention, to pragmatic considerations — the very issues on which a person's ethical commitment founders in the tempest of a personal crisis. But as my second example, let me take something less philosophical, more specific — the understanding of Communism. Is there a more important issue? Is there a worse-taught issue? Is there an issue on which we need information more desperately? Are there searching questions aimed at discovering true understanding rather than slogan memorizing? Clearly not. Here is a case where the label on the package will pass the educators ("teaching about Communism" is an acceptable goal), but the only sane way to do it (use Communist documents and speakers, [live, taped or filmed] as well as critical commentaries) is entirely unacceptable. The same applies to homosexuality, adultery, prostitution, violence, abortion, pornography, etc. — in short, to most of the topics that are likely to produce a personal moral crisis for the graduate of, or pupil in, our schools and that can be thoroughly and helpfully discussed there. Instead, they have to be discussed by the walking wounded in later life, too late for primary prophylaxis.

The second major weakness in the NAEP approach lies in the conceptualization of the goals and objectives. Without detracting from the very considerable merits of ETS and AIR, who did the work, the goals and objectives leave a great deal to be desired and bear the heavy signs of committee authorship. A few examples from the citizenship goals will indicate the kind of problem that exists.

Goal A is "Show concern for the welfare and dignity of others."[5] Of course, *showing* concern is not what we want; we want *having* concern. It's attractive to go for the "behavioral objective" formulation, but it focuses on external signs when we want something much deeper. Someone who does not show concern but who gets the ambulance is better than someone who weeps hysterically.

Objective G-1 is "Try to inform themselves on socially important matters and to understand alternative viewpoints."[6] Is the goal *trying?* Or is the goal *succeeding?* Suppose you find that everyone in the United States K—12 system is trying to inform themselves about something, but — e.g., because of incompetent teachers — failing dismally. Would you feel that citizenship/social studies education was succeeding? This is not a semantic issue. I suspect that, in some feeble sense, most people "try to understand" the use of bloody and destructive violence by political revolutionaries in this country. I think most of them (would say they) fail. I think that shows something about the gross inadequacies of citizenship/social studies education, not something about its success. They know nothing of the philosophy of violence; they could identify none or at most one of the half-dozen powerful reasons for the use of violence; it does not even occur to them that their own country was founded on violence and has perpetrated and institutionalized violence to a massive extent. They are examples of the *failure* of citizenship/social studies education, and an evaluation should so identify them.

The next major failure of the Assessment's effort lies in *interpretation*, and it really falls under two subheadings: interpretations by staff and interpretations by consultants whose report was published by National Assessment.

Here's an example of absurdly poor staff interpretation: "One indication that students do weigh alternatives rationally was seen in the group participation exercises; 67-79% at all three school ages gave a reason for a particular point of view at least once during the one half hour task."'

Giving a reason may be aimed at persuading others or rationalizing one's own decision; hence it is simply improper to take it as an indication of rational deciding. That error shows a very serious lack of understanding of what rationality is, and that lack of understanding shows up frequently.

What would be evidence of critical ability and of rational decision making? A case where prior prejudice won't give the right answer, where the answer must come by inference from the given facts of the case, in short a new problem case. None occur under Goal G "Approach Civic Decisions Rationally."

There's a pervasive overoptimistic bias in the interpretations. Why should one be inclined to think that young Americans' critical ability is anything less than ludicrous when a majority of 9-year-olds and a quarter of 13-year-olds think that a newspaper *can't be wrong?** That's after six of seven years of schooling.

Interpreting the global significance of the results was left to an advisory panel. I will indicate my interpretation of one small part and you'll see why I think the truly horrifying implications have not generally been recognized. Even with the data at hand, despite the many deficiencies already indicated, much more can be inferred than either staff or advisers have recognized. The conclusions are not both precise *and* highly probable. But policy decisions, contrary to the usual position, do not require these conditions. We operate off probabilities and possibilities, when the risk of not doing so is high; and in this area, that's surely the situation.

Let's take the respect for freedom of speech. It's often mentioned that 75% or more of the 13-year-olds thought that *no one on radio or television should be allowed to say either* that "Russia is better than the U.S." or that "Some races of people are better than others" or that "It is not necessary to believe in God."

What isn't so often said (though National Assessment staff noted it) is that 94% drew the line at *one or more* of these statements as a permissible media utterance (i.e., only 6% thought all were utterable). The 17-year-olds still show almost 80% refusing to allow all three, and the young adults still show 68% standing four-square against freedom of speech *in these medium-controversy examples*. When asked why they thought these statements *should* be allowed, only two thirds of the most stalwart (adult) sample could think of freedom of speech or ideas, etc., as a justification. One should perhaps quote as the most significant statistic the 76% of the adults who failed this simple test, treated as a simple recognition test of a well-known principle. Now how many of the remaining 24%, if on the board of a broadcasting station, would actually stick to their verbal endorsement of this principle? The evidence (from Hartshorne and May on) suggests that it will be far fewer — perhaps only 10% instead of 24%, perhaps only 2%. And are these examples extreme tests? On the contrary. Suppose the third quote was not "It is not necessary to believe in God" but "Belief in God is a sign of weak-mindedness and the source of most war and cruelty." Would we really have 20% left to count on?

Remembering that huge gap between professed moral principles and actual practice, how should we feel about a test of professed tolerance of other races under very mild stress as in A4 — "being willing to have someone from another race be your dentist or doctor, live next door to you, represent you in an elected office, sit at the next table in a crowded restaurant, stay in the same motel or hotel" — when we find that *43% of all age levels* draw the line at one or more of these possibilities. When it comes to the day when the respondent's daughter actually wants to date interracially, one can have little confidence that half of that 43% will remain with us (and I'd have to say that 10% would be a surprise).

Is it not disastrous that less than a quarter of young adults (22%) could give even one reason for and one reason against education deferments for the draft?

Now I would also say that *most* of the remaining questions are routine questions about routine behavior and knowledge, and the subjects performed routinely on them. One can draw little joy or sorrow from those other responses. But on the issues that test the capacity for crisis handling in the citizenship domain — although the tests are weak and the inferences from test performances to real performance very shaky — the results I have quoted represent most of the questions asked (since there were very few), and surely they represent significant features of the answers.

What did the Panel of Reviewers think of these results?" (Remember that their reactions represent the *only* evaluative global synthesis effort by NAEP).

By and large, they thought the results were pretty encouraging. A black panelist (Tobe Johnson) rightly complained about the "WASP" standards built into some questions. Larry Metcalf saw the same point, and some other biases, and cautioned us not to blame or credit the schools for the results.

But no one expressed horror at the plain ignorance and prejudice revealed here, and several expressed gratification. Evaluation results sometimes call for horror, and these ones do. As to blaming the schools, why not? There's no reason to think the schools couldn't change these results around if they tried and there's every reason to think they should try. No doubt families, communities and media are also to blame and would also resist the effort to change. That doesn't show it can't be done, and if it can be and should be and isn't, then those who don't do it must share the blame. Communities can be changed by their schools; schools aren't petrified by communities in law though they may be in fact.

So I'd sum up my reactions to the National Assessment effort as involving *grave* weaknesses of design and interpretation, as well as great technical virtuosity in many dimensions.

## Comments on the Taylor Chapter

Much of Taylor's excellent review is unexceptionable. I will just mention one disagreement.

Taylor says: "The assessment movement is counter to the humanizing movement in American education. It is promoting a closed rather than an open approach to curriculum." I think this is a very serious misconception. To expect schools to provide certain core learnings is not to inhibit their room for all sorts of innovation. To expect students to test well on understanding democracy is hardly inhibiting humanization!

## Conclusion

I have tried to develop new perspectives on the evaluation of citizenship/social studies, partly by describing new models and partly by criticizing the present ones. I hope this will lead us towards *more useful* evaluation and *more effective* education in this area. Nothing in our national priorities is more important.

## *NOTES*

1. See Appendix D, Reference 1, p. 3.
2. See Appendix D, Reference 14, p. 2.
3. See Appendix D, Reference 14, p. 2.
4. See Appendix D, Reference 14, p. 10.
5. See Appendix D, Reference 1, p. 9.
6. See Appendix D, Reference 1, p. 33.
7. See Appendix D, Reference 5, p. 93.
8. See Appendix D, Reference 5, p. 103.
9. See Appendix D, Reference 5, pp. 34-35.
10. See Appendix D, Reference 6.

# CHAPTER 5

# CRITIQUE OF NAEP OBJECTIVES AND PROCEDURES: CITIZENSHIP AND SOCIAL STUDIES

A. Guy Larkins
University of Georgia

The purpose of this chapter is to help social studies educators determine the extent to which the findings of the National Assessment of Educational Progress (NAEP) are grounded in sound practice. The following pages deal primarily with two topics — the objectives that NAEP used as a basis for developing test items (see Appendix A) and the procedures used to improve the quality of those items. They also contain less-detailed comments about sampling and data analysis.

## Objectives

One of the difficult tasks facing National Assessment was to decide which educational achievements ought to be assessed. NAEP's approach to this problem was to develop a set of objectives for each of the 10 learning areas. Test items, which were to have content validity for those objectives, were then written. The quality of test items in citizenship and social studies is dependent in part, therefore, on the quality of the objectives. That topic is discussed below in terms of two questions:

1. Do the objectives meet NAEP criteria?
2. Do they meet National Council for the Social Studies (NCSS) guidelines?

### *NAEP Criteria*

It is reasonable to expect NAEP objectives in citizenship and social studies to meet criteria established by NAEP. Those criteria are:

1. Specialists in the learning area must consider the objectives authentic from the viewpoint of the discipline. Scientists must agree the science objectives are authentic; mathematicians must agree upon the authenticity of the mathematics objectives, etc.
2. School people must recognize them as desirable goals . . . which schools are actively striving to achieve.
3. Parents and others interested in education must agree the objectives are important . . . .[1]

*Academic review.* The first criterion may reflect a time-honored assumption that school subjects should be based on parent academic disciplines; but citizenship has no definitive parent discipline(s). A number of scholarly areas are relevant to what ought to be taught under citizenship, but no discipline or group of disciplines stands in relation to citizenship the way that mathematics stands to school math.

Another problem is that more than any other Assessment area, citizenship is concerned with beliefs, attitudes and actions that require ethical justification. Scholars in a discipline cannot tell us, as a natural outgrowth of their academic training, what a student needs to know to be a good citizen in the ethical sense of the term good. Similar arguments could be made for social studies, especially to the extent that it includes citizenship education.

Having commented on some limitations of using scholars to judge the appropriateness of objectives, we now ask: How well was this criterion applied by NAEP? To what extent were scholars from the various social sciences and history included in the development of social studies objectives?

According to Frances S. Berdie, a NAEP representative:

> In the summer of 1965, 11 social scientists met for two and one-half days with members of the ETS staff to define the proper domain of an inquiry into the achievements of American education in this subject area.[2].

Apparently, however, Berdie used the term "social scientist" loosely. Of the 11 members of that committee, only 1 was explicitly identified as an active scholar in those disciplines that National Assessment claims are most relevant to social studies — history, geography, economics, political science, anthropology, sociology and social psychology.[3] At a later review conference, only 2 out of 19 participants were clearly identified with academic disciplines.[4] Although Berdie claimed that all 4 participants in a third review conference were social scientists,[5] they were each members of departments of education in various universities.[6]

Similar comments seem to be justified concerning the participation of academic scholars in the development of citizenship objectives.[7]

*Taught in schools.* NAEP's second criterion for objectives is that schools must be actively striving to achieve them. It is important to stress that this criterion has been consistently and repeatedly emphasized in the project's publications.[8] Therefore, it is surprising to find that the contracting agency responsible for developing citizenship objectives explicitly rejected that standard,[9] and that the rejection is stated in the midst of three separate claims by National Assessment that the criterion is important.[10] The obvious inconsistency is neither acknowledged nor explained. Nevertheless, the majority of NAEP citizenship objectives appears to be among those that schools are striving to achieve, but some are not.[11]

Is it also true that some social studies objectives fall outside the usual domain of the schools? The answer appears to be "no," but only if those objectives are considered one at a time. If the total set of social studies objectives is taken as approximately descriptive of school social studies curricula, children would receive an incredible dose of history and the social sciences. Far more content is implied by those objectives than schools can hope to teach.[12]

*Lay review.* Compared to usual practice, National Assessment made an earnest effort to meet the criterion that parents and others must agree that objectives are important. Despite the assumption that American public schools are answerable to citizens, assessments of educational achievement generally do not involve laymen in determining what ought to be measured. In contrast, NAEP held a series of regional lay conferences to review objectives, and major revisions to the social studies objectives resulted from those conferences.[13]

Could NAEP have done better? Yes. Persons who attended the lay conferences were nominated by groups such as the American Federation of Labor, National PTA, the United States Chamber of Commerce and the National Association for the Advancement of Colored People. This selection procedure nearly guaranteed overrepresentation of middle-class viewpoints.

*Summary.* How well did NAEP meet its own criteria for objectives? The answer is mixed: (1) Few people who are clearly identifiable as scholars in relevant academic disciplines were listed as reviewers of social studies and citizenship objectives. (2) The contracting agency for the citizenship objectives rejected the criterion that they must be goals that schools are striving to achieve, and the social studies objectives imply far more content than schools seriously attempt to teach. (3) The criterion of lay review was taken seriously and achieved reasonably well, but greater diversity of social-class representation on lay panels may be desirable.

The following section deals with the question: How well do NAEP objectives meet guidelines of the National Council for the Social Studies?

## NCSS Guidelines

A position paper titled *Social Studies Curriculum Guidelines* was published by the National Council for the Social Studies in the December 1971 issue of *Social Education.* The structure of most of what follows is taken from the rationale portion of that paper. It is divided into four subsections: Knowledge, Abilities, Valuing and Social Participation. The NCSS position on each of these topics is compared to the positions expressed in NAEP citizenship and social studies objectives.

*Knowledge.* NCSS guidelines and NAEP social studies objectives differ on the manner in which social studies should be drawn from history and the social sciences. NCSS appears to define social studies primarily as citizenship education. NAEP social studies objectives appear to define social studies primarily as condensed replicas of history and the social sciences.[14] NCSS explicitly rejects that definition.[15] Its position seems to be that those portions of the curriculum that are selected from history and the social sciences should be chosen for their relevance to the resolution of social problems.[16]

National Assessment citizenship objectives are generally consistent with the NCSS position on the relation of academic disciplines to social studies. Judicious selection from the disciplines, rather than an attempt to replicate them, seems to have been the rule. Citizenship objectives I F, I G, II and V are typical of those requiring that knowledge from the social sciences and history be applied to social problems;[17] but NAEP also includes several objectives that are outside the scope of the curriculum implied by the NCSS guidelines. For instance, NAEP citizenship objectives deal with such diverse topics as health and safety,[18] family relations,[19]social etiquette [20] and vocational education.[21]

*Abilities.* NCSS stresses divergent thinking, data processing and human-relations competencies.[22] One National Assessment social studies objective states that children should know that divergent thinking is useful to scientists, but it does not point to the need to foster divergent thinking among students.[23] Another objective, however, is consistent with the NCSS view on data-processing skills such as locating, organizaing and assessing data and source material.[24] Of the three human relations competencies cited by NCSS — sensitivity to others, communication skills and ability to cope with conflict and authority — the first and part of the third are mentioned by National Assessment.[25]

NAEP citizenship objectives attend to data-processing skills but are not as strong as social studies objectives in that area. Citizenship objectives are stronger than social studies objectives in human-relations skills,[26] but divergent-creative thinking is neglected. Furthermore, despite statements about the importance of dissent,[27] which implies diversity of opinion, the general tone of the citizenship assessment presses for conformity in values and in the stances taken on political-ethical issues.

*Values.* NCSS guidelines and both assessments agree that values should be dealt with in the school curriculum. Despite some ambiguity, NCSS opposes indoctrination of even basic values such as those contained in the Bill of Rights.[28]NAEP social studies objectives on values include the phrase "reasoned commitment" and, therefore, appear to be closer to the NCSS position than do the citizenship objectives.[29] Although one citizenship subobjectives states that rights are not absolutes and that they frequently conflict with each other,[30] other objectives present values as unqualified standards of proper behavior. In the majority of objectives that focus on values or on substantive social issues, the emphasis is on whether students take the "correct" stance, rather than on whether they make a rational choice.[31] That emphasis is out of harmony with the spirit of the NCSS guidelines.

*Social participation.* The central thrust of the NCSS position on social participation is absent from NAEP social studies objectives. That thrust includes: "Extensive involvement by students of all ages in the activities of their community . . . .[32] Those activities range from political campaigns to making important instructional and curriculum decisions in the school.[33]

Compared to NAEP social studies objectives, far more attention is given to social participation in the citizenship assessment. It is difficult to find a citizenship objective that does not include involvement by students in social problems.

They are based, however, on the assumption that correct social participation can be identified and ought to be expected.[34] In contrast, the NCSS position is more consistent with assessing the reasoning behind the positions taken by students, rather than assessing whether those positions fit majority expectations.

### Postscript to Objectives

The first half of this chapter has examined the question: How well do NAEP objectives meet NAEP criteria and NCSS guidelines? Again the answer is mixed. Some NAEP criteria were met better than others, and some objectives met NCSS guidelines better than others. In the final analysis, however, the critical question is: How well do NAEP objectives fit your assumptions about the proper content for social studies? It is this writer's opinion that those objectives will completely satisfy few people. But neither are the NCSS guidelines likely to completely satisfy most members of our profession. However, teachers who read and compare those objectives and guidelines will find many important statements with which they agree and many others that will stimulate them to clarify their opinions concerning appropriate goals for social studies.

## Exercise Development

### Introduction

Perhaps the most crucial task facing National Assessment was the development of valid measures of achievement. The criteria and procedures used by the project to develop citizenship and social studies exercises are examined in the present section of this chapter. Whether those exercises are valid is discussed elsewhere in this report. Among the criteria used by NAEP to judge the quality of their assessment instruments were those having to do with offensiveness, content validity, clarity and difficulty level. Among the procedures used to improve exercises were: (1) reviews by laymen, subject-matter specialists, technical advisors and the United States Office of Education (USOE), and (2) various field trials.

### Lay Reviews

One of the innovative features of National Assessment is that laymen were asked to review exercises to help insure that the assessment instruments would be acceptable to the general public. Although this is a useful and important way to involve laymen, it may have serious consequences for the content validity of exercises. For instance, it might not be possible to test some parts of a legitimate social studies or citizenship topic without getting into controversial problems. In those cases, the offensiveness reviews might cause exercises to be so modified that a decrease in offensiveness is accompanied by a decrease in validity. The potential danger to the validity of citizenship and social studies exercises is illustrated by the following selection of a few of the topics which the lay panels found offensive:

> References to specific minority groups should be eliminated whenever possible . . .
>
> Any reference to . . . the FBI, the President, Communism and specific organizations such as the Ku Klux Klan and labor unions might make an exercise offensive unless extreme care was used in the wording . . . .
>
> Exercises which show national heroes in an uncomplimentary fashion though factually accurate are offensive.
>
> Exercises which might be interpreted as putting the police or other authorities in an unfavorable light are offensive.[35]

The above examples were chosen because they seem to have more potential than others for reducing content validity. Examples could also have been selected that might improve validity. For instance, the lay panels objected to questions that violated the privacy of families and that expressed ethnocentric views.[36]

It is impossible to judge the extent to which the lay reviews affected content validity. But the potential impact was considerable. According to Finley and Berdie, "Citizenship exercises were reviewed and revised and re-reviewed so often that no figure [on the number revised as a result of the lay conferences] is meaningful . . . ."[37] Each of the five lay conferences examined citizenship exercises, and three of the five examined social studies. Only literature came close to receiving as much attention as the two areas most relevant to social studies educators. Our profession, therefore, should be particularly interested in the possible impact of the lay panels on content validity.

Other reviews that were relevant to the criterion of offensiveness include: (1) a lay conference that focused on whether exercises were trivial; (2) a review by the NAEP Technical Advisory Committee and (3) a review by USOE. Results of the USOE review illustrate that the criterion of offensiveness can have special impact on our area of the curriculum. Three of 4 exercises that were dropped, and all 11 exercises that were modified, were in citizenship.

### Subject-Matter Reviews

Following the lay reviews, exercises were sent to subject-matter specialists nominated by professional organizations such as the National Council for the Social Studies. Although this mailed review proved helpful in identifying problems with content validity, as few as two reviewers were used for each learning area It was impossible for reviewers to determine whether the larger content of an assessed learning area was properly sampled because each reviewer received only those objectives upon which the exercises he critiqued were based. Therefore, National Assessment decided to hold a series of subject-matter review conferences.[38]

At least four subject-matter conferences were held to review exercises for the social studies or citizenship assessments. Conferences were also held to produce additional exercises for 13-year-olds and for young people from minority groups.

An impression left by this series of conferences is that NAEP was determined to produce acceptable, appropriate and valid exercises. When the mailed reviews were not adequate, conferences were held to allow reviewers to meet face to face. When early conferences indicated a need for greater attention to assessing the achievement of minority group students, additional writing and review conferences were held. After each review, questionable exercises were returned to the contractors for modification. In some cases, new exercises were produced, which were also reviewed by subject-matter specialists. In terms of sheer number of reviews and revisions, it is difficult to imagine how the project could have given greater attention to face validity without violating their budget and production schedule. However, National Assessment may have been able to improve the reviews by using a more systematic and thorough within-conference procedure such as that described by Hunkins.[40]

### Field Studies

The content validity of exercises was determined primarily through the subject-matter reviews described above. Several field studies, however, focused on criteria that are relevant to the quality of test items, such as clarity and difficulty level.

The Assessment specified that each exercise must be written to meet one of three difficulty levels — very easy, moderately difficult or very difficult. An easy item is one that can be correctly answered by approximately 90 percent of the students.[41] One study attempted to determine whether "easy" items met the 90 percent criterion. The answer was "no." Exercise writers missed the mark by a wide margin.[42] Therefore, NAEP took special pains to produce additional easy exercises. Success in meeting the other difficulty levels was not determined.

Because high difficulty resulting from lack of clarity in any item was not desirable, three feasibility studies used interviews to spot problems in understanding test instructions, format, vocabulary and vague or ambiguous terms.[43] The importance of determining whether students understand test questions is dramatized by one of the results of a study that used low-achieving children: "At the 9-year-old level, the three students were such poor readers that each exercise had to be read aloud before they were able to answer."[44]

Two studies that focused in part on whether changes in format affect the difficulty of exercises were characterized by high internal validity.[45] The major conclusions of these studies were: open-ended questions tend to be more difficult than multiple-choice; the difficulty of multiple-choice exercises can be manipulated by changing the distractors; and including "I don't know" tends to reduce the number of correct responses to multiple-choice items. Since NAEP uses both multiple-choice and open-ended exercises, these results mean that caution should be used when comparing the results to different exercises. What may appear to be a difference in knowledge may simply be a difference in the difficulty of the item formats.

After exercises had been extensively revised as a result of the lay reviews, subject-matter reviews and initial field studies, final tryouts were held prior to the selection of items for inclusion in the actual assessment.[46] Exercises that were to be individually administered were tried out by interviewing six persons per item. Exercises that were to be group-administered were tried out by testing classroom-size sets of students.

The use of interviews in the tryout of individually administered exercises allowed for direct assessment of the clarity and difficulty of test items, but an indirect approach was used with the group-administered exercises. The classroom teacher and a representative of the contracting agency responsible for the tryouts each completed an observation form, which contained categories for such inappropriate student behaviors as apparent inattention or boredom. The interview procedure seems to be superior for detecting problems that are relevant to the validity of exercises.

### Summary

This author is impressed with National Assessment's emphasis on producing exercises that students can understand. The number of reviews, revisions, field studies and additional revisions indicates that this criterion was taken seriously. Nevertheless, NAEP seems to be operating on two questionable assumptions about how to improve exercise clarity.

The first assumption is that items that are clear to low-achieving students will also be clear to more able ones.[47] The opposite may be true, for instance, in cases where bright students are aware of the multiple meanings of ambiguous terms.

A second questionable assumption is that experts can make adequate judgments about whether an exercise will be understandable to children. The existence of field trials does not negate this assumption. Although National Assessment conducted several studies that were relevant to clarity, no field trial contained all of the following features: (1) interviews to uncover communication difficulties, (2) inclusions of all exercises that might be used in the final assessment and (3) adequate samples of respondents of various ages, abilities and backgrounds.

Despite the above criticisms, clarity is an important criterion, which NAEP tried diligently to meet.

In this writer's opinion, the criterion of three levels of difficulty is not as important as clarity. Even so, the criterion is a good one, which helped the Assessment avoid the narrow vision of assessing only a mid-range of achievements. The concern for difficulty levels also had an unanticipated benefit: failure to meet the 90 percent criterion forced a closer examination of the clarity of exercises.

### Reliability

An unusual feature of NAEP exercises is that standard estimates of reliability are inappropriate. Readers who are accustomed to seeing reliability reported as coefficients, such as .84 or .91, may be puzzled by the lack of such statements in Assessment reports. Types of reliability, however, that yield coefficients cannot be computed for single-item tests. An alternative is to report standard errors. Although this approach may be unfamiliar to many readers, they can depend on NAEP reports to be cautious in the narrative descriptions of findings. The language used to discuss findings is carefully chosen to reflect the amount of trust that can be placed on their reliability.

## Sampling and Data Analysis

In educational research, samples are frequently composed of local volunteers such as a few social studies classes in schools that are willing to cooperate. In contrast, National Assessment uses careful and thorough procedures that combine randomization with multistage cluster sampling. Readers of NAEP reports can be reasonably certain that the sample selected for each group in the assessment, such as 9-year-olds, is similar to the national population for that group.

Readers can also be reasonably certain that the procedures used to analyze Assessment data are appropriate. The technical competence of the NAEP staff and advisory committees is impressive. The cautious language used to report the technical aspects of the Assessment leaves the impression that they are aware of the proper applications and limitations of the procedures used to analyze assessment data.

## Conclusion

Despite the several criticisms in this article, it is the overall impression of this writer that NAEP used reasonable procedures. The technical aspects of the project appear to be sound. Most of the innovations, such as using laymen to review exercises, appear to be useful. Although a project as massive as National Assessment is bound to run into difficulties, even some failures, there are few research efforts in education of this scope and quality.

### *NOTES*

1. See Appendix D, Reference 7, p. 12.
2. See Appendix D, Reference 14, p. 4.
3. See Appendix D, Reference 14, p. 9.
4. See Appendix D, Reference 14, pp. 31-32.
5. See Appendix D, Reference 14, p. 7.
6. See Appendix D, Reference 14, p. 33.
7. See Appendix D, Reference 1, pp. 47-50.
8. See Appendix D, Reference 7, p. 5; Reference 1, pp. 2-3; Reference 14, pp. 2, 5, 7-8; and *Conferences on Development of Instrumentation for Assessment of the Progress of Education in the United States, June 15, 1965-June 30, 1967* (Denver: National Assessment of Educational Progress), p. 4.
9. See Appendix D, Reference 1, p. 5.
10. See Appendix D, Reference 1, pp. 2-3, 8.
11. See Appendix D, Reference 1, pp. 43-45.
12. See Appendix D, Reference 14, pp. 13-25.
13. See Appendix D, Reference 14, pp. 5-6.
14. See Appendix D, Reference 14, p. 9.
15. *Social Studies Curriculum Guidelines* (Washington, D.C.: National Council for the Social Studies, 1971), p. 857.
16. Ibid., p. 875.
17. See Appendix D, Reference 1, pp. 13-15, 21-26, 31.
18. See Appendix D, Reference 1, p. 11.
19. See Appendix D, Reference 1, pp. 43-46.
20. See Appendix D, Reference 1, p. 10.
21. See Appendix D, Reference 1, pp. 41-42.
22. *Social Studies Curriculum Guidelines* (Washington, D.C.: National Council for the Social Studies, 1971), p. 857.
23. See Appendix D, Reference 14, p. 12.
24. See Appendix D, Reference 14, pp. 11-12.
25. See Appendix D, Reference 14, pp. 9-10.
26. See Appendix D, Reference 1, pp. 30-31, 35-38.
27. See Appendix D, Reference 1, pp. 30-31.
28. *Social Studies Curriculum Guidelines* (Washington, D.C.: National Council for the Social Studies, 1971), pp. 859-861.
29. See Appendix D, Reference 14, pp. 26-27.
30. See Appendix D, Reference 1, p. 15.
31. See Appendix D, Reference 1, pp. 13-14, 19.
32. *Social Studies Curriculum Guidelines* (Washington, D.C.: National Council for the Social Studies, 1971), p. 859.
33. Ibid., p. 860.
34. See Appendix D, Reference 1, pp. 19-20.
35. See Appendix D, Reference 7, pp. 42-47.
36. See Appendix D, Reference 7, pp. 42, 46.
37. See Appendix D, Reference 7, p. 40.
38. See Appendix D, Reference 7, pp. 98, 103, 107.
39. See Appendix D, Reference 7, pp. 51-52.
40. F.P. Hunkins, *Validity of Social Studies and Citizenship Exercises* (Task II-Final Report, University of Washington, 1973), pp. 15-18.
41. See Appendix D, Reference 7, pp. 28-30.
42. See Appendix D, Reference 7, pp. 55-60.
43. See Appendix D, Reference 7, pp. 60-66.
44. See Appendix D, Reference 7, p. 73.
45. See Appendix D, Reference 7, pp. 77-78, 81-82.
46. See Appendix D, Reference 7, pp. 87-97.
47. See Appendix D, Reference 7, p. 61.

# CHAPTER 6

# VALIDITY OF SOCIAL STUDIES AND CITIZENSHIP EXERCISES

Francis P. Hunkins

University of Washington

The prime purpose of Task 3 was to determine the content validity of the social studies and citizenship exercises developed by the National Assessment of Educational Progress (NAEP). A secondary purpose of the investigation was to indicate if the exercises as developed could be utilized as models by teachers in developing their own evaluation instruments. Related aspects were considered as well by this investigator: the cognitive and affective levels of the exercises, the format of the exercises, the manner in which the exercises were administered and the age levels to which the exercises were geared.

Content validity is the most crucial criterion of any test exercise, for it appraises whether the exercise assesses what it is assumed to be measuring. Content validity centers on the representativeness or sampling adequacy of the content of the items. Such validity is basically concerned with the question, "Is the substance or content of the item depictive of the content or the universe of content being measured?"[1] Specifically, the question relating to the NAEP exercises is "Are these exercises representative of the objectives as developed by the National Assessment of Educational Progress?" Are the exercises doing what they are purported to do in relation to the objectives?"

## Social Studies, Citizenship — Goals and Objectives of National Assessment

Since content validity is judged in relation to goals and objectives delineated, the reader is provided with an abbreviated listing of the objectives in order to interpret this chapter more completely as well as to formulate some judgments relating to the value of the overall thrust of assessment in this area.

### Social Studies Objectives

Social studies is that area of the school curriculum that seeks to communicate about man in society. It is a shorthand term for such subjects as history, geography, economics, political sciences, anthropology, sociology and social psychology.

I. Have curiosity about human affairs.
II. Use analytic-scientific procedures effectively.
III. Are sensitive to creative-intuitive methods of explaining the human condition.
IV. Have knowledge relevant to the major ideas and concerns of social scientists.
V. Have a reasoned commitment to the values that sustain a free society.[2]

Each of the above objectives had several subdivisions with descriptions as to what knowledge and/or behaviors should be possessed or evidenced by individuals at various age levels. Some of these objectives had lengthy discussions of subunderstandings or subbehaviors. In determining the content validity for each item, the investigator and three judges scrutinized these descriptions under the major objectives.

### Citizenship Objectives

The National Assessment of Educational Progress did not create a precise definition of citizenship; however, its interpretations of citizenship are evident through the stated objectives.

I. Show concern for the welfare and dignity of others.
II. Support the rights and freedoms of all individuals.
III. Help maintain law and order.
IV. Know the main structure and functions of our government.
V. Seek community improvement through active, democratic participation.
VI. Understand problems of international relations.
VII. Support rationality in communication, thought and action on social problems.
VIII. Take responsibility for own personal development and obligations.

IX. Help and respect their own families (ages 9, 13, 17).

X. Nurture the development of their children as future citizens (adults).

## Specific Nature of the Task

As previously indicated, the Task 3 investigator was to review independently the NAEP social studies and citizenship exercises in order to inform NAEP, school personnel and the general public as to the extent to which they can place confidence in the exercises. Basically, the investigator analyzed exercises having one of four possible formats: completion, writing, multiple-choice and combination. The completion exercises required the student to fill in a blank or blanks or to provide a short answer following some stimulus or to arrange alternatives in correct locations. The writing responses required a student to write a sentence(s) or paragraph(s) or list responses. The multiple-choice exercises asked the individual to select from two or more alternatives, and the combination question requested the individual to select from one or more alternatives and then to respond by completion or writing.

### *Procedure for Processing the Exercises*

The investigator was assisted by three panel members in processing the exercises to estimate whether they measured the objectives intended and to record descriptor information on each exercise dealing with age level, format, manner administered and the like. Data were recorded on a format sheet identical to the following one:

---

Format for Recording Exercise Data

---

Learning area:
Item:
Item format:
Individual or group administered:
Age:
Objective:
Cognitive-affective levels:[4]
Content validity:
Context validity:
Model for teacher:
Age appropriateness:
Other:

---

The sequential instructions followed by the investigator and his assistants for processing both the social studies and citizenship exercises were identical:

1. Read each exercise and record descriptor information.
2. Make judgments as to whether the exercises had content validity, context validity, exemplary qualities for teacher use and age appropriateness.
3. Read appended responses to each exercise if present.
4. As a safeguard against misjudgment reconsider the content-validity question with regard to how individuals actually responded.

Judges, responding independently to all exercises, initially agreed on the content validity and exemplary qualities of a surprising 95% of the exercises. Then judges and the major investigator as a group discussed their judgments of the content validity and exemplary qualities as well as context validity, cognitive/affective levels and age appropriateness. In cases of disagreement on content validity and exemplary qualities reasons for differences were contemplated. The group had little difficulty in reaching complete agreement on content validity and exemplary qualities. The investigator put aside for additional deliberation those exercises where consensus was not achieved on other categories of secondary interest to this report and later made the final judgments on such matters.

## The Results

Results are reported for released and unreleased social studies exercises and for released and unreleased citizenship exercises. (Readers may be interested to know that the panel found the unreleased exercises in both social studies and citizenship similar to the released exercises.)

### *Social Studies Exercises: Released and Unreleased*

Table 1 summarizes the data relating to all social studies exercises, both released and unreleased. The table records information relating to the total number of items for each major objective as well as the number of exercises for each subobjective. Since this chapter lists only the major objectives, individuals wishing information about the specific subobjectives should refer to Appendix A.

Perhaps the most significant point is that of the 194 exercises prepared for social studies, 85 percent (164) of them were considered valid by the investigator and the panel members. This should enable us to have some degree of faith in the exercises as truthful in measuring what they state they are measuring.

Just looking at the overall percentage of exercises can be misleading. Not all objectives are represented by equal numbers of exercises, and also the exercises emphasize different cognitive and affective levels.

Objectives II and IV had significantly more exercises prepared — 51 and 83 exercises, respectively. Whether this suggests a hierarchy of importance regarding the objectives remains to be seen. Perhaps this loading is due to the objectives themselves: Objective II relates to whether individuals use analytic-scientific procedures. With the emphasis on process today, perhaps this distribution of exercises reflects current curriculum status. However, this inference is somewhat suspect, for if process were the emphasis, the exercises would not cluster at the lower cognitive levels. Perhaps the number of exercises is related to the number of subdivisions in the objectives; for example, Objective IV has six subdivisions. Or perhaps the number of exercises reflects simply the technical difficulties of creating feasible exercises for some subobjectives. Moreover, all exercises were reviewed by panels of lay persons, subject-matter specialists and the United States Office of Educaiton (USOE). Since social studies — and citizenship — are often controversial, some exercises were likely to be ruled out in this review process. But if NAEP has stated that social studies learning should be related to five major objectives, then all dimensions of these objectives should be presented adequately by exercises. That no exercises exist for Objectives IIB, IIF, IIIC and VD and that only one exercise exists for Objectives IIIA and IVA makes it impossible to assess whether individuals are demonstrating behaviors couched in these objectives. National Assessment needs either to eliminate these subobjectives or to create exercises for them.

**TABLE 1. Summary of Social Studies Released and Unreleased Exercises**

| Objective | # of items | Item Format | | | | Admin. | | Age Levels | | | | Cog. Lev. | | | | | | Affec. Lev. | | | | | # Valid | # Not Valid | # Exem. | # Con. Val. | Age App. | Totals |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Comp. | Wri. | Ml. Ch. | Comb. | Ind. | Group | 9 | 13 | 17 | Ad | K | C | A | AN | S | E | Re | Rs | Va | Or | Ch | | | | | | |
| IA | 10 | | 4 | | 6 | 10 | – | 6 | 5 | 2 | 3 | 3 | 6 | | | | | 3 | 3 | | | | 9 | 1 | 10 | 10 | 10 | Exercises<br>for I: 14<br>7% |
| IB | 2 | | | | 2 | 2 | | | 2 | 2 | 2 | | 1 | | | | 1 | 1 | | | 1 | | 1 | 1 | 1 | 1 | 1 | |
| IC | 2 | | 1 | | 1 | 2 | | 1 | 1 | 1 | | | 2 | | | | | | 1 | 1 | | | 2 | | 2 | 2 | 2 | |
| IIA | 5 | | 2 | 3 | | 2 | 3 | 1 | 2 | 3 | 3 | | 1 | | 3 | | | 1 | | | | | 4 | 1 | 3 | 3 | 4 | |
| IIB | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | |
| IIC | 26 | 8 | 4 | 14 | | 9 | 17 | 9 | 15 | 12 | 10 | 11 | 8 | 4 | | 1 | 1 | | 1 | | | | 25 | 1 | 24 | 24 | 24 | |
| IID | 9 | | | 9 | | | 9 | 7 | 5 | 4 | 4 | 1 | 5 | | 1 | | 2 | 1 | | | | | 6 | 3 | 7 | 7 | 8 | Exercises<br>for II: 51<br>26% |
| IIE | 11 | | 8 | 3 | | 3 | 8 | 7 | 5 | 4 | 4 | | 2 | | 8 | | | | | | 1 | | 10 | 1 | 11 | 10 | 11 | |
| IIF | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | |
| IIIA | 1 | | 1 | | | | 1 | | 1 | 1 | 1 | 1 | | | | | | | | | | | 1 | | 1 | 1 | | |
| IIIB | 6 | | 5 | 1 | | 2 | 4 | 2 | 6 | 6 | 5 | | | | 6 | | | | 1 | | 2 | 1 | 6 | | 6 | 6 | 6 | Exercises<br>for III: 13<br>7% |
| IIIC | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | |
| IIID | 6 | 1 | | 5 | | | 6 | 4 | 5 | 5 | 3 | | 5 | | 1 | | | | | 1 | | | 6 | | 6 | 6 | 6 | |
| IVA | 1 | | | 1 | | | 1 | | 1 | 1 | 1 | | 1 | | | | | | | | | | 1 | | 1 | 1 | 1 | |
| IVB | 4 | | 2 | 2 | | 1 | 3 | 1 | 1 | 2 | | 1 | 2 | 1 | | | | | | | | | 3 | 1 | 3 | 2 | 2 | |
| IVC | 14 | | 8 | 6 | | | 14 | 4 | 7 | 10 | 10 | 4 | 7 | 3 | | | | 1 | | | | | 9 | 5 | 9 | 10 | 10 | |
| IVD | 18 | | 4 | 14 | | 3 | 15 | 14 | 7 | 6 | 5 | 13 | 1 | 1 | 2 | | | 2 | | | 1 | 1 | 13 | 5 | 13 | 13 | 13 | |
| IVE | 27 | | 3 | 24 | | 4 | 23 | 12 | 18 | 19 | 20 | 13 | 14 | | 2 | | | 1 | 1 | | 1 | | 19 | 8 | 18 | 18 | 18 | Exercises<br>for IV: 83<br>43% |
| IVF | 19 | | 1 | 18 | | 1 | 18 | 4 | 12 | 15 | 14 | 11 | 7 | 1 | | | | | | | | | 16 | 3 | 16 | 16 | 16 | |
| VA | 9 | | 3 | 1 | 5 | 5 | 4 | 5 | 6 | 7 | 6 | 3 | 2 | | | 3 | | | | 5 | | 3 | 9 | | 9 | 9 | 9 | |
| VB | 11 | | 1 | 2 | 8 | 5 | 6 | 1 | 4 | 11 | 10 | | 1 | | 1 | 1 | 4 | | | 1 | | 9 | 11 | | 10 | 10 | 10 | |
| VC | 3 | | | | 3 | | 2 | 1 | | 2 | 3 | 2 | | | | | | | | 1 | | 2 | 3 | | 3 | 3 | 3 | |
| VD | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | |
| VE | 7 | | 4 | | 5 | 6 | 1 | 3 | 2 | 2 | 2 | | | | 3 | 2 | | | | 2 | 3 | 2 | 7 | | 7 | 7 | 7 | Exercises<br>for V: 33<br>17% |
| VF | 3 | | | | 3 | 1 | 2 | 1 | 2 | 2 | 2 | | | | | 1 | | | | | | 2 | 3 | | 3 | 3 | 3 | |
| Grand Totals | 194 | | | | | | | | | | | | | | | | | | | | | | 164 | 30 | 163 | | | 194 |

**TABLE 2. Summary of Citizenship**
**Released and Unreleased Exercises**

| Objective | # of items | Item Format | | | | Admin. | | Age Levels | | | | Cog. Lev. | | | | | | Affec. Lev. | | | | | # Valid | # Not Valid | # Exem. | # Con. Val. | Age App. | Totals |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Comp. | Wri. | Ml. Ch. | Comb. | Ind. | Group | 9 | 13 | 17 | Ad | K | C | A | AN | S | E | Re | Rs | Va | Or | Ch | | | | | | |
| IA | 9 | | 1 | 4 | 4 | 7 | 2 | 1 | 9 | 6 | 5 | 2 | 2 | | | | | | | | 5 | | 6 | 3 | 6 | 6 | 6 | Exercises for I: 22<br>14% |
| IB | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | |
| IC | 4 | 1 | | | 3 | 3 | 1 | | 1 | | 2 | 2 | | | | | | | 1 | | | 1 | 3 | 1 | 2 | 2 | 2 | |
| ID | 7 | 5 | | | 2 | 7 | | 2 | 6 | 2 | 4 | 3 | 2 | | | | | | | 1 | 1 | | 2 | 5 | 2 | 2 | 2 | |
| IE | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | |
| IF | 1 | | | | 1 | 1 | | | | | 1 | | | | | | | | | 1 | | | 1 | | 1 | 1 | 1 | |
| IG | 1 | | | | 1 | 1 | | | | 1 | 1 | 1 | | | | | | | | | | | | 1 | | | | |
| IIA | 3 | | | | 3 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | | | | 1 | | | | | 1 | 1 | 2 | 1 | 2 | 2 | 2 | Exercises for II: 11<br>7% |
| IIB | 6 | | 2 | 3 | 1 | 5 | 1 | 2 | 5 | 3 | 4 | 3 | 2 | 1 | | | | | | | | | 4 | 2 | 4 | 4 | 4 | |
| IIC | 2 | | | | 2 | 2 | | | | 1 | 1 | | 2 | | | | | | | 1 | 1 | | 2 | | 2 | 2 | 2 | |
| IIIA | 4 | | 3 | | 1 | 4 | | 1 | 1 | 2 | 3 | | 4 | | | | | | | | | | 4 | | 4 | 4 | 4 | Exercises for III: 8<br>6% |
| IIIB | 2 | | 2 | | | 2 | | 1 | | 1 | 1 | | 1 | | | 1 | | | | | 1 | | 2 | | 2 | 2 | 2 | |
| IIIC | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | |
| IIID | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | |
| IIIE | 1 | | | | 1 | 1 | | | | | 1 | 1 | | | | | | | | | 1 | | 1 | | 1 | 1 | 1 | |
| IIIF | 1 | | | 1 | | 1 | | | | | 1 | | 1 | | | | | | | | | | 1 | | 1 | 1 | 1 | |
| IVA | 1 | | 1 | | | 1 | | 1 | 1 | 1 | | | 1 | | | | | | | | | | 1 | | 1 | 1 | 1 | Exercises for IV: 46<br>30% |
| IVB | 23 | 7 | 4 | 10 | 2 | 15 | 8 | 8 | 12 | 12 | 10 | 17 | 5 | | 1 | | | | | | | | 9 | 14 | 7 | 7 | 7 | |
| IVC | 6 | 1 | 2 | 3 | | 2 | 4 | | 3 | 5 | 6 | 2 | 4 | | | | | | | | | | 4 | 2 | 3 | 3 | 3 | |
| IVD | 14 | 2 | 6 | 1 | 5 | 12 | 2 | 2 | 10 | 9 | 8 | 6 | 4 | | 1 | | | | | | 1 | | 8 | 6 | 6 | 6 | 6 | |
| IVE | 2 | 1 | | | 1 | 2 | | | 2 | 1 | 2 | 2 | | | | | | | | | | | 2 | | 1 | 1 | 1 | |
| IVF | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | |

### TABLE 2. Summary of Citizenship Released and Unreleased Exercises (Continued)

| | | Item Format | | | | Admin. | | Age Levels | | | | Cog. Lev. | | | | | | Affec. Lev. | | | | | # Valid | # Not Valid | # Exem. | # Con. Val. | Age App. | Totals |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Objective | # of items | Comp. | Wri. | Ml. Ch. | Comb. | Ind. | Group | 9 | 13 | 17 | Ad | K | C | A | AN | S | E | Re | Rs | Va | Or | Ch | | | | | | |
| VA | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | |
| VB | 3 | | 1 | 2 | | 1 | 2 | | | 2 | 3 | | 1 | | 1 | 1 | | | | | | 1 | 2 | 1 | 2 | 2 | 2 | |
| VC | 10 | | | 1 | 9 | 9 | 1 | 1 | 1 | 5 | 8 | 1 | 2 | | | 1 | | | | | 2 | 7 | 8 | 2 | 7 | 8 | 8 | Exercises<br>for V: 17<br>11% |
| VD | 1 | | | | 1 | 1 | | | | 1 | 1 | | | | | | | | | | | 1 | 1 | | 1 | 1 | 1 | |
| VE | 3 | | | 1 | 2 | | 3 | | 1 | 1 | | | | 1 | 2 | | | | | | | 1 | 3 | | 3 | 3 | 3 | |
| VF | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | |
| VIA | 9 | | 7 | 1 | 1 | 8 | 1 | 5 | 5 | 4 | 4 | 2 | 4 | 1 | | 2 | | | | | 2 | 1 | 8 | 1 | 8 | 8 | 8 | Exercises for VI: 9<br>6% |
| VIB | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | |
| VIIA | 12 | | 5 | 3 | 1 | 9 | 3 | 4 | 6 | 7 | 8 | 3 | 5 | | 2 | | | | | 1 | 2 | | 3 | 9 | 2 | 3 | 3 | |
| VIIB | 2 | | | | 2 | 1 | 1 | 2 | 1 | | 1 | | 1 | | | | 1 | | | | | | 2 | | 2 | 2 | 2 | |
| VIIC | 1 | | 1 | | | 1 | | | 1 | | | | 1 | | | | | | | | | | 1 | | 1 | 1 | 1 | |
| VIID | 3 | | 1 | 2 | | 3 | | | | 1 | 2 | 1 | 1 | 1 | | | | | | | | | 2 | 1 | 2 | 2 | 2 | Exercises<br>for VII: 20<br>13% |
| VIIE | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | |
| VIIF | 2 | | 2 | | | 2 | | 2 | 2 | 1 | 1 | | 2 | | | | | | | | | | 1 | 1 | 1 | 1 | 1 | |
| VIIIA | 8 | | | 1 | 7 | 8 | | 4 | 1 | 2 | 2 | 3 | 1 | | | | | | | 3 | 2 | 1 | 6 | 2 | 6 | 6 | 6 | |
| VIIIB | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | Exercises<br>for VIII: 9<br>6% |
| VIIIC | 1 | | | | 1 | 1 | | | 1 | | | | 1 | | | | | | | | | | 1 | | 1 | 1 | 1 | |
| VIIID | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | |
| IXA | 5 | 3 | | 1 | 1 | 5 | | 1 | 1 | | 3 | 1 | | | | | | | 1 | | 2 | 1 | 2 | 3 | 2 | 2 | 2 | Exercises<br>for IX: 10<br>7% |
| IXB | 5 | | | | 5 | 5 | | 2 | 2 | | 1 | 2 | | | | | | | | | 1 | 1 | 1 | 4 | 1 | 1 | 1 | |
| IXC | 0 | | | | | | | | | | | | | | | | | | | | | | X | X | | | | |
| Grand Totals | 152 | | | | | | | | | | | | | | | | | | | | | | 93 | 59 | 84 | | | 152 |

### *Citizenship Exercises: Released and Unreleased*

Table 2 reports a summary of data relating to all of the citizenship exercises. Skimming the table allows one to see which objectives were stressed, which were ignored regarding exercise representation and the content validity of these exercises. As is true with the social studies exercises, the citizenship exercises, both released and unreleased, favor certain objectives. Objective I, "Show concern for the welfare and dignity of others," had a total of 22 exercises while Objective IV had 46 exercises. The emphasis on Objective IV, which deals with the main structure and function of our government, might suggest that citizenship as conceived by the exercise developers, is primarily political. However, the overall objectives do suggest a broader interpretation of citizenship. Moreover, several citizenship subobjectives have no exercises at all.

It would be important to know whether the number of exercises per objective represents a conscious plan or is the result of just how the exercises happened to fall as developed and approved. This disproportionate weighting of some objectives is a problem that needs attention.

Not only did the investigator and the panel members find these gaps in exercise existence, but they judged only 93 (or 61 percent) of the total 152 exercises developed as having content validity. One needs to exercise extreme caution in assessing the level of citizenship functioning of individuals when 39 percent of the exercises are not valid. Even fewer were considered exemplary — only 55 percent.

## Final Comments

As a whole, the panel concurred with the investigator that the social studies exercises as a total group, both released and unreleased, were superior to the citizenship exercises, both released and unreleased; a greater percentage (85 percent) of social studies exercises possessed content validity as opposed to only 61 percent of the citizenship exercises. Also, a greater number of social studies exercises were deemed exemplary.

Interpreting the results of National Assessment is tied up with whether one favors or opposes any national assessment. However, if we assume that the reader is in favor of national assessment, and if we assume that the majority of the exercises are valid, it is still difficult to determine precisely whether the schools are to be praised or blamed for the current levels of understanding and functioning of individuals in the area of social studies and citizenship. Certainly, the schools are not the only institutions in society that educate. Therefore, National Assessment can tell us what the levels of understanding, skills, attitudes and so on are of various groups in our population, but it cannot tell us that these levels are entirely the results of good or poor schooling. But schools do need feedback even if of a general nature. The Task 3 investigation points up some needed adjustments in certain exercises and objectives to afford educators even more reliable input for their decision making about what schools themselves need to do to enable individuals to become truly effective persons, citizens for the present and future.

### *NOTES*

1. Fred V. Kerlinger, *Foundations of Behavioral Research* (New York: Holt, Rinehart and Winston, 1965), p. 446.
2. See Appendix D, Reference 14, pp. 9-27.
3. See Appendix D, Reference 1, pp. 7-39.
4. Cognitive and affective levels were defined in two sources. Benjamin S. Bloom, *A Taxonomy of Educational Objectives: The Cognitive Domain* (New York: David McKay, 1956), and David Krathwol, *A Taxonomy of Educational Objectives: The Affective Domain* (New York: David McKay, 1956).

# CHAPTER 7

# A RATING OF SOCIAL STUDIES EXERCISES BY SOCIAL STUDIES EDUCATORS

June Chapin
College of Notre Dame
Belmont, California

When it comes to interpreting the findings of assessments in citizenship and social studies, thoughtful people must ask whether actual performance levels are adequate. The National Assessment of Educational Progress (NAEP) aims to report; it does not aim to congratulate or deplore. Judgments are left to others.

One task of the National Coucil for the Social Studies (NCSS) review, then, was tackling the hard question of satisfactory performance levels on social studies exercises. (Unfortunately, neither time nor resources were available for asking like questions about citizenship exercises.) The results, when all was said and done, came from exercises. Although social studies exercises were, indeed, related to objectives, every exercise was also expected to have some significance of its own, to stand on its own. Consequently, some examination of the worth of the exercises themselves seemed needed. Moreover, since the exercises were written, NCSS had published its *Social Studies Curriculum Guidelines*.[1] It seemed sensible to ask whether the exercises were in line with these guidelines.

A panel of nine social studies educators was accordingly selected. (Their names along with members of other panels appear elsewhere in this report.) In addition, the author of this chapter was designated chairperson with responsibility for designing rating sheets and processing data. Members of the panel came from different geographic areas. They were comparatively young — seven were under 45 and three under 30 years of age. The male-female ratio was about equal. About a third were members of minority groups, important in the light of questions about the suitability of exercises for such groups. Professional backgrounds were varied, but social studies consultants and higher-education faculty were decidedly better represented than classroom teachers.

A training session was held using the NCSS guidelines and rating forms for each major matter to be examined. Members of the panel practiced on exercises from the citizenship assessment for independent ratings and for group discussion to clarify criteria. After the training session each rater proceeded independently. Exercises were identified only by code number, not by objective. Panel members agreed more often than not, but at times spread their responses across all the available categories. When disagreements in ratings did occur, they were reported out, not buried. Differences in ratings came from ambiguity in the NCSS curriculum guidelines, varying interpretations of the meaning of the exercises and outright differences in the judgments of the raters.

Since each panel member made literally over 1,000 ratings — many exercises had subparts — the ratings were processed by computer. Almost 3,000 pages were turned out. Although the full NCSS report includes complete data, space limitations require that summaries only and some too few illustrations of ratings of "realistically satisfactory" performance be included in this report.

## Compatibility with NCSS Curriculum Guidelines

National Assessment states: "Social Studies is that area of the school curriculum that seeks to communicate about man in society. It is a shorthand term for such subjects as history, geography, economics, political science, anthropology, sociology and social psychology. . . ."[2] At this point readers may want to refer to the social studies objectives listed in Appendix A. NAEP objectives, at least, appear to define the social studies as mainly a miniature version of the social sciences.

The NCSS *Social Studies Curriculum Guidelines* offer a somewhat different perspective. The guidelines see the social studies as essentially citizenship education — although their definition differs some from that of NAEP's citizenship objectives — and emphasizes social problems. The guidelines identify four, integrally related components of social studies education: (1) knowledge; (2) abilities in thinking, human relations and obtaining information; (3) valuing; and (4) social participation. These guidelines are to be viewed as both a stimulus and a guide to evaluate existing social studies curriculum and to work for improvement. They may be thought of more in terms of what ought to be, according to the professional organization, than what actually exists in schools. The curriculum guidelines may reflect more of the spirit of the times and more of the cutting edge. Nonetheless, considerable overlap appears between them and Assessment objectives.

One task of the panel was to point out more precisely the relationship between social studies exercises and the NCSS guidelines. The panel worked with the following question: *Into which of the main components of the NCSS guidelines do you think this exercise best fits? (1) knowledge, (2) abilities, (3) valuing, (4) social participation?*

The panel allotted (1) 46% of the exercises to knowledge, (2) 31% to abilities, (3) 19% to valuing and (4) 2% to social participation (and 2% to "no response").

Striking is the very small percentage placed in the "social participation" category. NAEP has separated assessment of citizenship and social studies. Had the panel rated the citizenship exercises, this category may well have increased. An examination of released citizenship exercises appears to support this claim.[3] However, the panel's ratings do give some indication of National Assessment's view of the social studies; social participation is not stressed.

According to the panel's rating, "abilities" plus "valuing" came to about half of the social studies exercises, and "knowledge" the other half. Many may be pleased that NAEP social studies exercises, in the judgment of the panel, give one fifth of their total to "valuing." An important question is whether the "knowledge" category is too high. Still, in practice, the testing programs of most schools probably devote a much higher percentage of test items to "knowledge." Many teachers and much of the public may be satisfied with the degree of attention given to the "knowledge" category in the exercises.

## Desirability of the Exercises

NAEP exercises might be balanced among the four components discussed above and still be considered of little value or hardly in keeping with guidelines recommendations. It seemed important that the panel give a formal and careful estimate of the general worth of each exercise on its own. As another task, then, the panel members rated by these directions: *"Using the NCSS guidelines as criteria, estimate the general desirability of the exercise: (1) little, (2) some, (3) moderate, (4) high or (5) very great."*

The panel rated the exercises as shown in Table 3.

**TABLE 3. Summary of NAEP Social Studies Exercises Classified by General Desirability**

| Degree of Desirability | Percentage of Exercises | Number of Observations |
|---|---|---|
| **Very great** | 26% | 879 |
| **High** | 40 | 1,356 |
| **Moderate** | 21 | 722 |
| **Some** | 6 | 194 |
| **Little** | 3 | 97 |
| **No response** | 4 | 118 |
| **Total** | 100% | 3,366 |

If "desirability" is defined as including the first three categories, "very great," "high" and "moderate," then approximately 87% of the exercises, by the panel's criteria of NCSS guidelines, are desirable. If the first two categories, "very great" and "high," are used, the panel judged two thirds desirable. Fewer than 10% of the social studies exercises were rated as of "some" or "little" worth, but NAEP might do well to look carefully at these exercises. (Readers may find it helpful to compare these ratings with those of Francis Hunkins' Task II panel in Chapter 6 of this report. That panel judged validity for NAEP objectives without reference to NCSS guidelines.)

## Realistically Satisfactory Performance

Schools have long used standardized tests. Ordinarily such tests were constructed to sort out the most able from the least able students — that is, the midscore was to be roughly "half correct" to allow the most and least able to spread themselves up and down the distribution scale. Since some benchmarks by which to interpret were needed, mean or median scores were obtained. Lack of better benchmarks than these meant that adequate performance was interpreted as the average of the population on whom the test was normed: it was "good" to be above the norm and "bad" to be below.

National Assessment, however, did not aim to distinguish the most or least able individuals but to find out what proportion of young people could or could not do important tasks. NAEP, therefore, moved away from norm-referenced tests and over to performance tests. Although NAEP has presented "national percentages of success" as a way of summarizing findings, these are not norms in the older sense. (The temptation to regard them so will be strong.) Moreover, 50% success can not necessarily be considered "adequate," nor 75% "good."

For example, 17-year-olds were offered the following exercise.

> Below is a discussion that was held in 1966. As you read it, try to decide what the two speakers primarily disagree about.
>
> Speaker I: The United States should fight a limited war in Vietnam while seeking a negotiated settlement. Winning of the war in itself won't do any good. The United States must aim instead at seeing that the South Vietnamese have improved education, democratic government, security of life, and then deal with poverty and the lack of medical care. Financial aid, advice and technological know-how are what are really needed, but . . . .
>
> Speaker II: Improving living conditions is a good idea, but our primary job is fighting. The United States can't permit itself to be pushed out of an area where it is committed. If we withdraw, we would be telling that part of the world threatened by Communist aggression that we either cannot or will not maintain our position. All that really matters is our power position in international affairs.
>
> What do the two speakers primarily disagree about?
>
> ☐ What power and poverty mean in international affairs.
>
> ☐ Whether the United States is actually capable of controlling South Vietnam by force.
>
> ☐ The extent to which the United States should be involved in Vietnam and the motives for its involvement.

☐ Whether Communist aggression in Vietnam is worse than a lowered standard of living in the United States.

☐ I don't know.

Seventy-five percent of the 17-year-olds were successful.[4] Is this "good" performance? (The panel thought "yes" and judged that 61-80% should be able to perform successfully.) Thirteen-year-olds were asked:

A. Do you think teenage students should help decide what courses will be offered in their school system?

__Yes __No __Undecided

B. Please give a reason for any answer you selected.

NAEP reports that 64% of the 13-year-olds gave acceptable reasons for their choice in part A.[5] Is this "good" performance? (The panel thought "yes" and judged that 61-80% should be able to perform successfully.)

Moreover, National Assessment exercises were written at very roughly three difficulty levels; on some exercises 90% were expected to respond correctly, on others 50% and on others only 10%. While considerable discussion could be given here to the issues inherent in writing exercises with expectations of 100% success, or "mastery," suffice it to say here that NAEP, for good reason, decided against exercises written with such expectations.

The problems of giving meaning to performance levels and of clarifying ideas about what is significant to assess seem clearer when attempts are made to say what satisfactory performance levels are. The panel's judgments are simply those of nine people, although competent people. Their ratings, however, may give some guidance.

The NCSS panel was asked this question: "Realistically what level of performance nationally for the age level being considered would satisfy you for this exercise? (1) less than 2% correct, (2) 20-40%, (3) 41-60%, (4) 61-80% or (5) more than 80%?" Notice that the panel's purpose was not to guess or predict an actual level of performance but to consider what a satisfactory level should be. Notice also that the panel aimed to be "realistic." It is easy to hope that everyone will be able to do "everything." In present education and social conditions and with all the variations among young people, defining "satisfactory" as "everyone should do it" is to dodge the problem.

Although a summary of ratings perhaps obscures more than it clarifies, a summary is made in Table 4.

**TABLE 4. Summary of NAEP Social Studies Exercises Classified by National Level of Realistically Satisfactory Performance**

| Realistically Satisfactory Performance Levels | Percent of Exercises | Number of Observations |
|---|---|---|
| **Less than 20%** | 2% | 79 |
| **Between 20 and 40%** | 3 | 186 |
| **Between 41 and 60%** | 13 | 673 |
| **Between 61 and 80%** | 34 | 1,858 |
| **Greater than 80%** | 48 | 2,611 |
| **Total** | 100% | 5,407 |

On the great bulk of social studies exercises the panel was "realistically satisfied" with proper responses from no less than 60% of the population. In almost half of the exercises the panel wanted proper responses from 80% of the population.

## Summary

The data presented here are summary data; they hide and gloss over many fine points. For example, on some single exercise the panel may have judged the exercise of high value in terms of desirability but was realistic enough to be satisfied with a fairly low percent of proper responses. Best interpretations are made from examining the panel's rating of each exercise. Such examination fits with NAEP's point of view that each exercise should be examined by itself.

With suitable limitations in mind, it appears from panel ratings that National Assessment exercises are closely related to the NCSS *Social Studies Curriculum Guidelines* (with the exception of the component, social participation) and possess some obvious degree of desirability according to panel judgments from the guidelines. Professionals would be realistically satisfied if 61% or more of the population gave correct responses on most social studies exercises.

---

*NOTES*

1. *Social Studies Curriculum Guidelines* (Washington, D.C.: National Council for the Social Studies, 1971).
2. See Appendix D, Reference 14.
3. See Appendix D, Reference 5.
4. See Appendix D, Reference 8, p. 14 (Exercise RS118). Released exercises will be published in full in succeeding NAEP reports.
5. See Appendix D, Reference 8, p. 52 (Exercise RABO2).

# CHAPTER 8

# AN ANALYSIS OF A SELECTED SET OF SOCIAL STUDIES EXERCISES: KNOWLEDGE OF INSTITUTIONS

C. Benjamin Cox
University of Illinois

## Introduction

One of the major tasks identified for the overall assessment of the national assessment of social studies by a researcher group from the National Council for the Social Studies (NCSS) was an interpretation of the data generated in the citizenship assessment of 1969-70 and the social studies assessment of 1971-72. The expectation was that ways could be found to compare the findings from the two assessments since they shared areas of interest. Perhaps a confirmation of citizenship findings or even some indication of growth between the two assessments would somehow fall out of the comparisons.

It was further expected that a number of different groupings of the data would allow some useful embellishments in our interpretations. In general, these groupings require the selection of items from the total assessment to be treated together. The going idioms at the National Assessment of Educational Progress (NAEP) for such a group of items are "color scheme" and "theme." The computer is asked to put the items together in order to compute all manner of statistical results, which presumably say more to the researcher about the population he is working with than if he had results only from individual items. The options in the NAEP model are individual item data and data from clusters of items or "color schemes."

The intention in the NCSS assessment was to use a propitious combination of these two options. A variety of color schemes was envisioned — e.g., factual items, value items, skill items, historical items, sociology items, economics items, etc.; items grouped according to the NCSS guidelines; items grouped by NAEP objectives in social studies; items grouped by a variety of content themes other than the mother disciplines, e.g., racial concerns, moral questions, civil rights, etc.; and items assessing critical thinking, logic, judgment and decision making, reasoning, etc.

There was also the intention of casting explanatory hypotheses in an effort to account for good and bad performances. There was no expectation that such hypotheses could be tested, of course, since National Assessment data are purely descriptive. This intention would be greatly enhanced by the data breakdown built into the NAEP treatment. The treatment involves the breakdown of data by ages 9, 13, 17 and adult; Northeast, Southeast, Central and West regions of the U.S.; two sexes; two races; four levels of parental education; and seven size and type of community (STOC) categories. Thus, the seeker of explanations is guided to ask, "Why do blacks perform worse than whites on this item about the Supreme Court?" When a cluster of items in a color scheme is under consideration, he may ask, "Why do persons in the upper socioeconomic areas of the cities know more about their democratic rights and duties than persons who live in rural areas?"

A further intention within this interpretive task was to draw implications for public policy. For example, one is tempted to infer from the consistently poor performances of blacks in this assessment that something is awry in the conduct of schools, in curriculuas, in teaching strategies, in the society, in the assessment procedures, in the assessment content or possibly in the black population itself. Some of these, at least, would be subject to alteration by changes in public policy.

A final intention was to translate findings and data into forms more useful to an assortment of users. While the project's treatment of data is not at the sophisticated statistical level of much of education research, it is several cuts above simple addition and subtraction. Multivariant analysis, regression formulas, correlation coefficients and chi squares are absent here; but there are sufficient NAEP idioms, such a raw p, delta p, cut-off, hinge and eighth as well as standard terms in the nomenclature, such as category, variable, standard error, median and mean, to boggle the mind. The craft of the interpreter, one would suppose, is to say obtuse things in ways that are meaningful to readers.

### *Some Limitations — Fortuitous and Otherwise*

We have suggested some of the limitations of this interpretation. Chief among these is the impossibility of finding correlations between a performance in one situation and a performance elsewhere. Educators are fond of discovering such relationships — e.g., between I.Q. and reading or belief in x and belief in y. The reason for this limitation is the sampling and testing schemes employed by National Assessment. As more elaborately explained in

Larkins' chapter, NAEP has constructed sample populations of some 27,000 persons in each age level of 9, 13 and 17, and 9,000 adults. However, out of deference for the physical and mental comfort of these persons, the total number of items being used to assess a learning area is divided into 10 to 14 sets or packages so that each package contains only 1/10 or 1/14 of the items. An individual or group selected for testing as a part of an age sample would receive only one package containing, say, 20 items out of a total of 200 to 300 in a learning area. Thus, the items contained in a given package would be answered by only 2,000 to 2,500 persons or so. Correlations between items within packages would be possible since these items would have been answered by the same persons. While NAEP has experimented with this further treatment of assessment data, the data presented here are not dealt with in this way. Correlations between items in different packages are not statistically permissible, however, because the two sets of items would have been answered by two different groups of people.

Statisticians substantiate that when appropriate sampling methods are used, clusters of items, as in a color scheme, may be treated as if they represent the performance of a total age level population. Thus, answers to a cluster of 10 items, while possibly the actual performances of 10 groups of 2,500 persons — each group having answered one item in the cluster — may legitimately represent the performance of the entire sample. An assumption of group equivalency is made in this regard, of course.

A second serious limitation relates to a series of fortuitous events that range temporally over a year and a half, geographically from Washington, D.C., to San Francisco, and ethically from principal to expediency. The upshot of the difficulties is that severed budgets, delayed computers and contract deadlines have restricted this interpretation to fewer than 50 items out of some 200 in the social studies assessment and to two color schemes out of a planned half dozen or so. We anticipate that in the future this interpretation will be completed with all social studies items included and with whatever additional color schemes are on the NAEP computer tape.

We hope to correct another limitation in that future expansion of this interpretation. This document contains only limited references to the earlier findings in the citizenship assessment. An unfortunate deterrent to any comparison of the results of the two assessments is the fixing of summary data — i.e., the color schemes or clusters — by different rubrics. A presumption is made, however, that some goals in the two areas are so similar that comparisons are possible and that some items in the two assessments can be found that appear to ask similar things.

*Items included.* This interpretation is limited to the consideration of the items displayed in Appendix C, where they are grouped into two color schemes. The two schemes, originally requested for this NCSS assessment in the categories of "Knowledge of Rights and Duties" and "Knowledge of Institutional Structures," largely parallel the social studies items used in NAEP's recent publication *Political Knowledge and Attitudes 1971-72.*

## Reporting the Data

All data, all statistical treatments of data and all interpretations of statistics contain inherent distortions and inaccuracies. The fact is that statistics, statisticians and interpreters of statistics impose their own peculiar frames on data. The danger occurs in the pretense of purity, which is not and cannot be. That pretense is not made about the data and interpretations reported here.

For the most part, *percentage of success* is the basic statistic used in the Assessment. That means simply the percentage of persons from a designated population who marked the item correctly or successfully, as judged by NAEP. This statistic is called *raw p* in NAEP jargon; we use raw p, percentage and percentage of success as equivalent terms.

Another National Assessment convention derived from two percentages of success is also used in this report. One of these values is the national percentage of success on a given exercise. This value always refers to the performance of all the persons in an age level — i.e., 9, 13, 17 or adult — who responded to an exercise, usually about 2,500 persons, give or take 200 to 250. There is never a combination of ages in this or any statistic in reporting NAEP data. The other percentage of success used refers to the performance of a group within the exercise population, e.g., males, females; blacks, whites; persons living within the Northeast, Southeast, Central or West region. The difference between the national percentage of success (raw p) and the group's percentage of success (raw p) is reported as *delta p.*

The values referred to thus far are illustrated in Table 5. The data are presented as representing the performance of 13-year-olds on exercise number 406011.

**TABLE 5. Sample Exercise Data — 13-year-old Respondents**

| | National | Region | | | | Sex | |
|---|---|---|---|---|---|---|---|
| | --- | Southeast | West | Central | Northeast | Male | Female |
| **Raw p** | 73.7% | 69.9% | 74.2% | 74.1% | 76.3% | 74.4% | 73.0% |
| **Delta p** | --- | −3.8 | 0.5 | 0.4 | 2.6 | 0.7 | −0.7 |
| **N-count** | 2,687 | 706 | 660 | 674 | 647 | 1,381 | 1,306 |

Thus, 73.7% of the 2,687 13-year-old respondents who marked this exercise, answered it correctly. Of the 706 13-year-old respondents who lived in the Southeast region, 69.9% answered the exercise correctly. The Southeastern 13-year-olds' percentage of success is 3.8 percentage points below the national percentage of success, and so on. Delta p-values will be used in several displays in this report when referring to individual exercises.

Another statistical arrangement is used when dealing with summary data, i.e., groups of exercises or color schemes. In this case, all the exercises in the color scheme are ordered from highest to lowest delta p-values of a group within an age level. The median value in this list of delta p's is taken as the typical performance of the group on that set of exercises. Thus, exercises in the list whose delta p-values are at or near this median value are said to be typical of this group's performance in the domain of the color scheme. A complex statistical operation is employed to determine, on the basis of such information as the standard error and the difficulty of each exercise for the group, which of the exercises in the color scheme were answered atypically by the group. Another set of statistical functions is performed to determine whether a given delta p-value is large enough to be called significantly or reliably different than the national raw p. Thus, when dealing with color schemes, the computer program gives to the interpreter a list of exercises ordered by the delta p-values for which a median has been determined, and in which performances atypical of the "group effect" have been demarcated and performances statistically significantly different than the national percentages have been flagged. These elements are boons to the interpreter since they give him something extra to say about the group's performance on the set of exercises, i.e., that is performed comparatively better or worse than expected (atypically) on the demarcated exercises relative to its performances on the other exercises and that it performed significantly better or worse than the whole age sample on the flagged exercises. Our use of the typical and significant nomenclature will remain consistent throughout the report, e.g., significant will always mean *statistically* significant.

An example of such a display of a color scheme is shown in Table 6.

**TABLE 6. Color-Scheme Display of Exercises and Performances**

**Age: 9 Size and Type of Community: Rural**
**Color Scheme: Knowledge of Institutions (Structures and Rights and Duties)**

| Delta P-Values | National Percentages of Success (Raw P) | Exercises: Short Text |
|---|---|---|
| 4.20* | 88.93% | Unreleased text |
| 3.03 | 35.92 | Which Job Health Dept. - Inspect. Restaurant |
| - - - - - - - - - - - | | |
| 0.33 | 83.18 | Unreleased text |
| (median: 0.220) | | |
| 0.11 | 73.82 | Who Responsible Fair Trial/Judge |
| - - - - - - - - - - - | | |
| −4.83 | 57.54 | Who Head Town Government/Mayor |
| −8.49* | 46.85 | Unreleased text |

In this short color scheme, rural 9-year-olds performed *typically* at a level very near the national raw p. The performances on the two exercises nearest the hypothetical median (medians must be computed in even-numbered lists in order to make a 50-50 split) are only .11 and .33 above the national raw p. A group's typical performances do not have to be close to the national raw p's, of course. The reference point for the judgment of typicality is the group's own median, the "most typical" performance in the color scheme. Typical performances are contained within calculated limits on either side of the group's median. It is possible for all of a group's performances to lie within the limits of typicality; ordinarily, however, at least a few extreme scores in a group's dispersion on a color scheme are judged to be atypical. The best performances in the example, in which the rural 9-year-olds are 4.20 and 3.03 percentage points above the national raw p's, are better than their typical performances. Likewise, their worst performances, 8.49 and 4.83 percentage points below the national raw p's, are worse than their typical performances.

Furthermore, the two extreme delta p-values of -8.49 and 4.20 are flagged by the computer as being significantly different than the national percentages for the exercises. The two delta p-values of -4.83 and 3.03 are merely chance variances from the national percentages. While statistically significantly different scores normally appear at the low and high ends of a dispersion, significance is a function of the size of the delta p score on an exercise, not of the dispersion *per se*. A group's performances on all the exercises in a color scheme could be significantly better or worse than the national performances on the exercises. Thus, a delta p-value may be typical and significant, typical and not significant, atypical and significant, or atypical and not significant.

The wily reader of statistics will have inferred by now that an assumption of relatedness of exercises must be made to support the assertions of typicality and atypicality in all cases. The existence of a valid color scheme is the foundation of all such group summaries. Statistically, the most valid color schemes would be produced by factor analysis. Then one could tell operationally which exercises go together. National Assessment has toyed with this notion, but color schemes to date, including the two involved here, have content validity only. That, too, could be enhanced with the help of a few like-minded experts. Hunkin's chapter offers a surfeit of visual judgments on which exercises in the social studies and citizenship assessments are valid candidates for their designated offices. Also Chapin's chapter reports jury decisions that relate to exercise validity. These materials were not available, however, when the judgments on these two color schemes were made. To the extent that this researcher's decisions effected these groupings, no appeal was made to a wider expertise. That lonely culpability notwithstanding, the schemes are not precisely as requested; so help came from somewhere.

A final comment on the presentation of group data in a color scheme concerns the dispersion of delta p-values on either side of the typifying median. Earlier we explained the Assessment's convention of using the median value in an ordered array of exercise performances as the most typical performance of a group. It was also suggested in that discussion that pegging the "group effect," "median value" or "typical performance" of, say, 13-year-old blacks at 11.22 percentage points below the national percentage of success for all 13-year-olds is a useful but incomplete bit of information. The median delta p tells us nothing about the dispersion of delta p-values above and below that median value.

The median value example cited above (delta p = −11.22%) will illustrate the usefulness of also noting the range or dispersion of delta p-values. In the color scheme involved in that instance, there are 26 exercises. For the entire listing of 26 exercises, the group effect (median delta p) of 13-year-old blacks is 11.22 percentage points below the national raw p. However, the poorest performance of this group on any exercise in the list was 28.98 percentage points below the national raw p. At the other end of the scale, the best performance for the group was on an exercise where it scored 7.52 percentage points above the national raw p. The median and range can be shown in some such display as the following:

| Lowest Delta P | Median Delta P | Highest Delta P |
|---|---|---|
| −28.98 | −11.22 | +7.52 |

While none of the exhibits in this chapter include this dispersion information, frequent and consistent references are made to the extreme limits of the distributions in the interpretive discussions. The highest and lowest delta p-values are often atypical in that they depart farthest from the median delta p.

### *Limitations of the Data*

What with all the complexities of treatment, one would presume that a wealth of hypotheses could be invented about youth, learning, schools and cultural effects in this country from NAEP data. Numbers of observations and complexities of treatment notwithstanding, the wealth available is extractable primarily through the loosest kind of inferring. In the statistical world, trade in loose inferences is bearish. Moreover, National Assessment, being a political creature, is constrained from engaging in such creative hypothesizing.

More specific to the point here, it is the nature of the data that poses the harshest limitations. Primarily what we have here are disjunctive data, to coin a phrase, rather than conjunctive data, to coin another; and they are disjunctive in several ways.

First, as explained earlier in this chapter, the grand sample of, say, 25,000 persons in an age level turns out in the real world to be a composite of, say, 10 samples of 2,500 persons each. An assessment in a learning area may be comprised of 250 exercises. However, no one person and no one sample takes that sort of test; instead, in the real world, each works with only 25 social studies exercises. Later they are quilted together to compose the social studies assessment. Given the right assumptions about sample randomness, test reliability, exercise validity, etc. and the right and properly rigorous statistical controls, this disjunctive approach can produce some reliable knowledge of a gross sort. This sample and test disjunction needs to be kept in mind nonetheless.

There is also a disjunction across ages that makes generalizations comparing one age level with another tenuous. The problem is mainly one of test equivalency, though all the sampling problems dealt with above are active here too. While the big goals and objectives in social studies are shared by all the age levels, common sense alone dictates different translations of these goals, different activities, different content, etc., at least to a degree. Thus, while all age levels may share some questions, each level may have a unique combination overall and some unique questions of its own. A simple display of the "Rights and Duties" color scheme (Table 7) utilized in this paper will illustrate the problem. The color scheme as a whole is comprised of 22 exercises, variously shared by the different age levels.

**TABLE 7. Exercise Distribution in Rights and Duties Color Scheme**

| **Exercise:** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Age 9** |  | x |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Age 13** |  | x | x |  |  |  |  |  | x |  | x | x | x | x | x |  |  |  | x | x |  | x |
| **Age 17** | x | x |  | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| **Adult** | x | x |  | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |  | x | x | x |

Cross-age conclusions between 17-year-olds and adults present the least difficulty; the color scheme for them is substantially equal. But cross-age inferences between ages 13 and 17 must be drawn from half the number of exercises, a dubious equation at best; and nothing useful can be said about 9-year-old respondents.

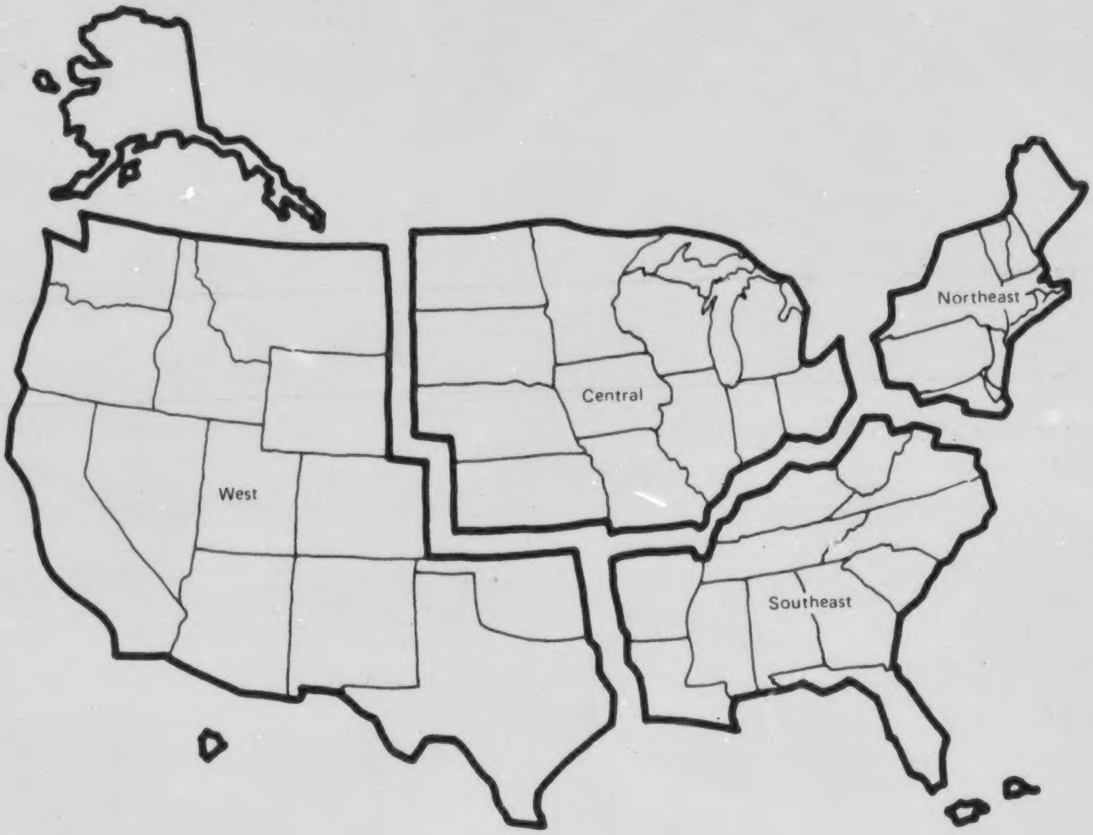
Though the data are superfluous to the point, we will display the exercise sharing for the other color scheme, "Knowledge of Structures" (Table 8), as a handy checkpoint for interested readers.

**TABLE 8. Exercise Distribution in Knowledge of Structures Color Scheme**

| **Exercise:** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Age 9** | x |  | x | x |  | x |  |  |  |  |  |  |  |  |  |  |  |  | x |  |  |  |  |  |  |  |
| **Age 13** |  | x |  |  | x |  | x | x | x |  |  |  |  |  | x | x | x | x |  | x | x | x | x | x | x | x |
| **Age 17** |  | x |  |  | x |  | x | x | x | x | x | x | x | x | x | x | x | x |  | x | x | x | x | x | x | x |
| **Adult** |  | x |  |  | x |  | x | x | x | x | x | x | x | x | x | x | x | x |  | x | x | x | x | x | x | x |

As we have shown in several ways, though without a formal layout, NAEP data are broken down for each age level in a variety of interesting and pertinent ways. As a starting point for this discussion, we will display these breakdowns in Exhibit 2.

### EXHIBIT 2. Categorical Breakdown of NAEP Assessment Data



**National**

**Region**

- Southeast
- West
- Central
- Northeast

**Sex**

- Male
- Female

**Race**

- Black
- White
- Other

**Parental Education**

- No high school
- Some high school
- Graduated high school
- Post high school
- Unknown

**Size and Type of Community (STOC)**

- Low metro
- Extreme rural
- Small place
- Medium city
- Main big city
- Urban fringe
- High metro

These categories, with the help of the inset map of the United States, are presumably self-explanatory with the exception of size and type of community (STOC). Because the definitions of these seven STOC categories are fairly technical and involved, the National Assessment designations follow. Readers are urged to check these definitions with some care since they represent a propitious blending of geographic and socioeconomic concepts. Please note that the STOC categories apply only to repondents in school at the time of the assessment.

*Low metro.* This category comprises 7.25% of the total sample and represents individuals attending schools whose students have the largest proportions of parents not regularly employed and/or on welfare.

*Extreme rural.* This category comprises 9.15% of the total sample and represents individuals attending schools whose students have the largest proportions of parents engaged in farm work.

*Small place.* This category comprises 28.29% of the total sample and represents individuals attending schools in communities with populations less than 25,000 and not classified under extreme rural.

*Medium city.* This category represents 17.40% of the total sample and represents individuals attending schools in communities with populations between 25,000 and 200,000 and not classified in low metro, extreme rural or high metro.

*Main big city.* This category comprises 8.41% of the total sample and represents individuals attending schools within the city limits of cities with populations greater than 200,000 and not classified in low metro or high metro.

*Urban fringe.* This category comprises 17.35% of the total sample and represents individuals attending schools in metropolitan areas served by cities with populations greater than 200,000 but outside the city limits and not classified in low metro, extreme rural or high metro.

*High metro.* This category comprises 12.15% of the total sample and represents individuals attending schools whose students have the largest proportions of parents in professional or managerial positions.

Now, back to our main point about the disjunctiveness of these data. Each of the age levels, 9, 13, 17, adult, is categorizable in each of the ways shown in the exhibit. Thus, we can show the percentage of success (raw p) of all 9-year-olds (national), for example, on any exercise designated for 9-year-olds and the group effect (median raw p) and dispersion of all 9-year-old performances across a color scheme of exercises. Further, we can show these same statistical performances (raw p, median raw p, etc.) for 9-year-olds who live in the Southeast, *or* who are males, *or* who are black, *or* whose parents have not gone to high school, *or* who live — more accurately, who go to school — in a low metro community. The repeated use of the disjunctive *or* in the sentence dramatizes the disjunctive data problem. We cannot show by these data the performance of Southeastern, black, male 9-year-olds whose parents have not gone to high school and who go to school in a low metro community. In fact no cross-category combinations of any kind are available in these data.

This is not a limitation of which NAEP is unaware. It has, in fact, contemplated future assessments in which some such combinations are entailed. On a national scale, however, the difficulties are awesome; on a regional scale, they are only horrendous. A major part of the problem relates to sample size and to the inferring of population facts from sample data. A hypothetical case will illustrate the point. Suppose there are 25,000 9-year-olds in the sample. Only 2,500 of these youngsters would be administered a given exercise. Perhaps 660 of this group would live in the Southeast and only one half of these, 330, would be males. In the Southeast perhaps one third, 110, would be black and certainly fewer than one half, 55, would have parents who never attended high school. Finally, possibly less than one fifth of these would be categorizable as low metro. That would provide a sample of a dozen or so, and possibly fewer, from which to infer population facts. The confidence level of one's statistical inferences would be low. Whether, for all the bother, it would improve on pure guess is open to speculation.

With this examination of the data-producing machinery and the nature of the data produced, we will turn to an examination of the data themselves. Our displays and discussions will make use of the concepts and information explored in this first part.

## A Review of the Data in the National Assessment of Social Studies, 1971-72

We begin our discussion of the results of the social studies assessment with reference to the three basic displays of group effect in Exhibits 3, 4 and 5. The values shown in these three exhibits are median raw p-values. In Exhibit 3 the group effect is shown for each age level in the categories of national; Southeast, Northeast, Central and West region; black and white race; high and low metro STOC; and no and post high school parental education over the entire set of exercises included in this report. These include 6 exercises for 9-year-olds, 26 exercises for 13-year-olds, 42 for 17-year-olds and 41 for adults. This total group of exercises is presumed to reflect the respondents' overall "Knowledge of Institutions."

Exhibit 4 summarizes the group effect for each age level in the same categories as in Exhibit 3 for the color scheme, "Knowledge of Institutional Structures." Exhibit 5 displays median raw p-values for the color scheme, "Knowledge of Institutional Rights and Duties." The results for the "Structure" theme are based on 5 exercises for 9-year-olds, 16 exercises for 13-year-olds, 21 exercises for both 17-years-olds and adults. The "Rights and Duties" theme results are based on 10 exercises for 13-year-olds, 21 for 17-year-olds and 20 for adults. No 9-year-old results are possible here because only one exercise was judged to treat this theme.

Generalizing from these data in Exhibits 3, 4 and 5 is a bit hazardous; but to be perfectly honest and open, one must say that for every display. Rather than repeating it endlessly, we will assume that all will proceed from this point with caution.

The striking feature of the three graphs is the general configuration of lines. If we could make an assumption of test equivalency across the four age levels — which, of course, we cannot — we could say that the years between 9 and 13 are something of a loss for young persons in the U.S. as far as this aspect of social studies is concerned, but the social and political world comes alive for them sometime after age 13. By age 17, then, most will have learned all they ever will; in their adult life as many will lose as will gain in knowledge of the political and social institutions that govern their lives and their rights and duties within them. While a number of constraints will not allow us to base so strong a statement on these data alone, we suspect the hints are indeed there and something like this is in fact the case. Even discounting the

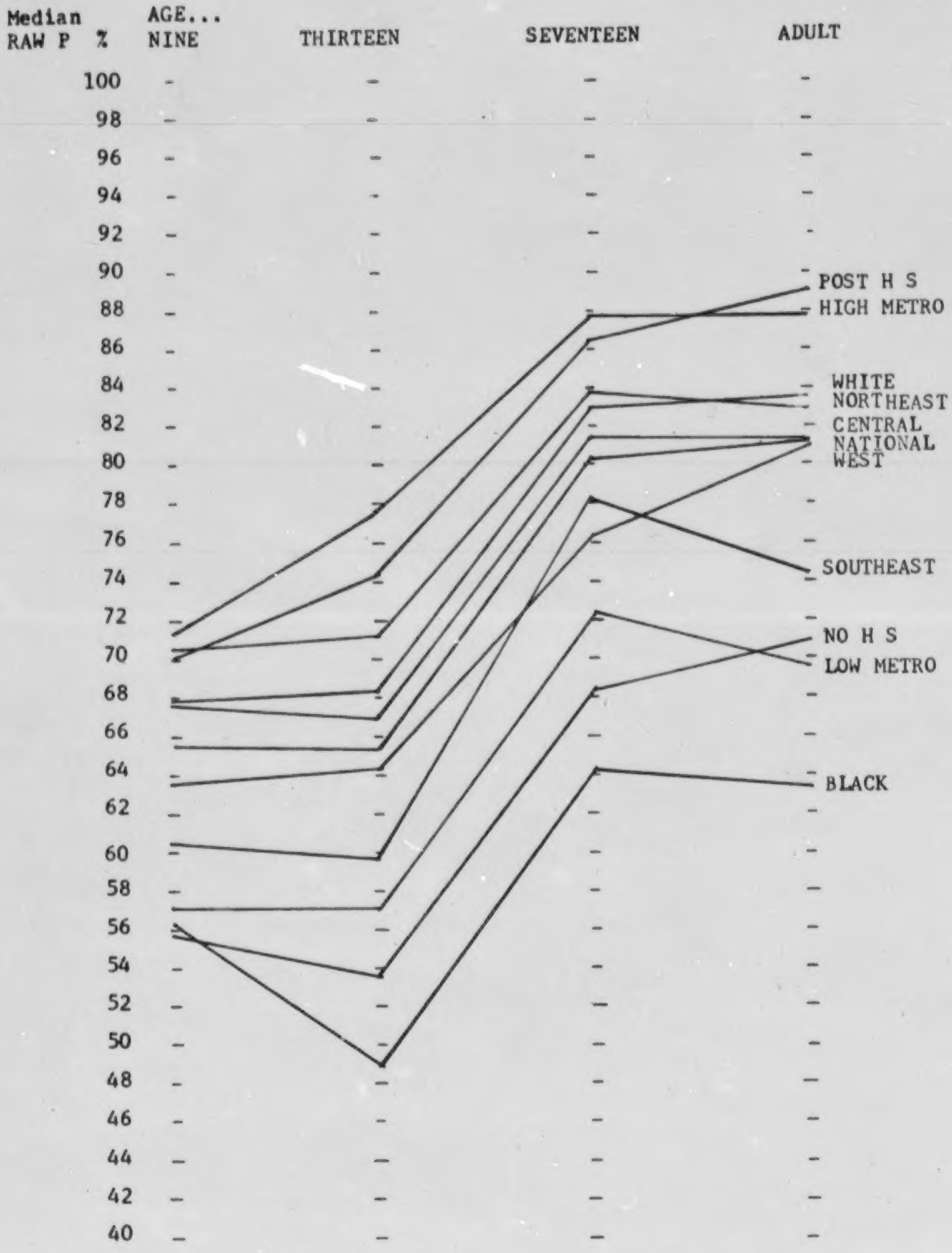


EXHIBIT 3. Knowledge of Institutions - Median Raw P-Values - Nation, Region, Race, STOC, Parental Education - Social Studies 1971-72

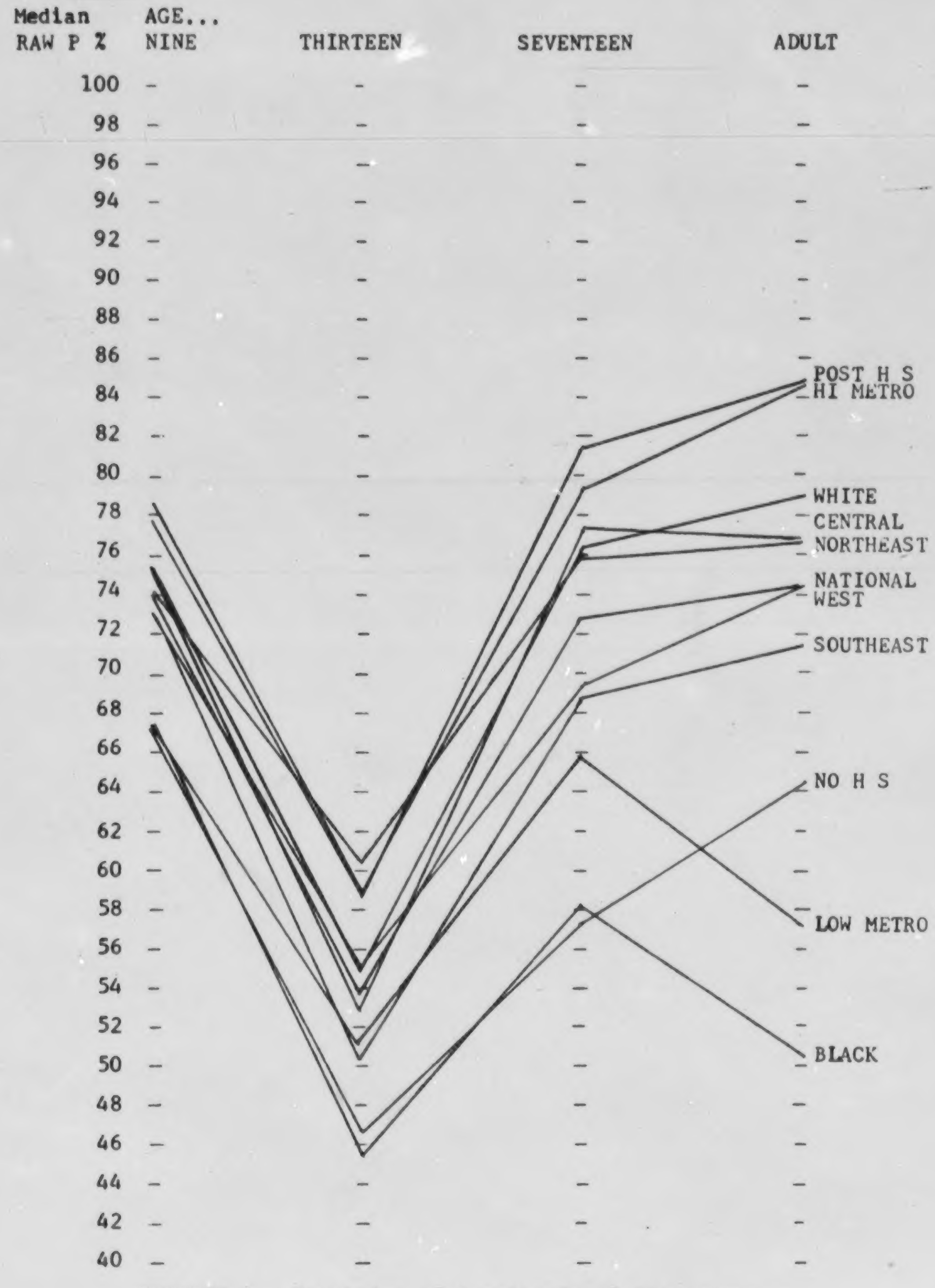


EXHIBIT 4. Knowledge of Institutional Structures - Median Raw P-Values - Nation, Region, Race, STOC, Parental Education - Social Studies 1971-72

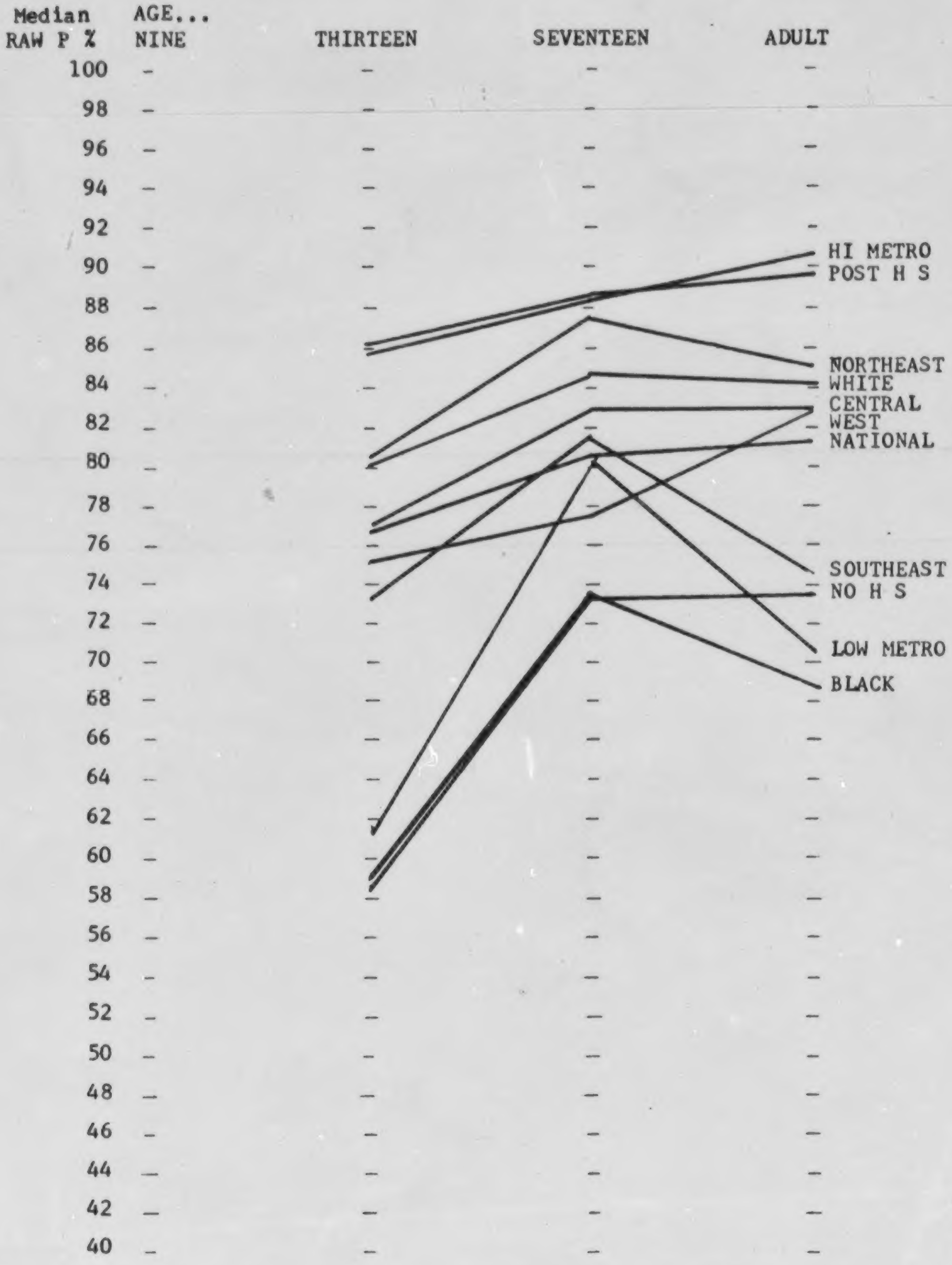


EXHIBIT 5. Knowledge of Institutional Rights and Duties - Median Raw P-Values - Nation, Region, Race, STOC, Parental Education - Social Studies 1971-72

results at age 9, which are probably spurious due to the few exercises, we are left with the dramatic rise from age 13 to age 17 and the noticeable cooling off after age 17. The sets of exercises given to 13-year-olds and 17-year-olds are only marginally comparable, but there is near equivalency between the 17-year-old and adult exercises. A similar, but less pronounced, rising and cooling effect is evident in the earlier citizenship assessment for some groups, principally males, medium size cities, urban fringes and smaller places, on a similar theme.[1]

So, by rounding, squinting and allowing in several ways, we can probably say that something is fairly successful in increasing young persons' knowledge of institutions by as much as 25%, in relative terms, between ages 13 and 17 (an absolute gain of about 15 percentage points in these data). We may also be able to say that schools in general and social studies in particular play some part in this growth. Most students have almost two years of U.S. history and one year of government between ages 13 and 17. That does not suggest that there are grounds for pride in this rough calculation, however. A good part of the increase may be accounted for by Piagetian hypotheses concerning the general shift from the concrete operations stage to the formal operations stage of cognitive development in this age range. Understanding of this content and interest in it may be as much products of development as schooling.

We infer the effect of schooling on what apparently happens to persons when they leave school. They retain, at least in part, what they learned as teen-agers 10 years or so earlier; but they do not continue to grow. Now, of course, a statement like that is not warranted by these data. We know absolutely nothing about these particular young adults when they were 17 except by inference and conjecture.

They were born between, say, 1937 and 1947 and half or more would have graduated from high school between 1954 and 1964 at about 17 years of age. Schools were not much different than from 1971-72 when this assessment was made. Furthermore, these young adults knew something of World War II, or at least its aftermath, and lived through Korea and Viet Nam; some served in the armed forces there or elsewhere. They saw their president and his assassin killed on television and watched the whole panorama of strife and dissent throughout the sanguine 60s. Some would have voted a dozen times perhaps.

Therein lies a tragedy that these data suggest. Learning in this country, at least as it measured by these exercises and as it concerns this content, stops at the schoolhouse door, not, as some have insisted, on the way in, but on the way out. Schoolmen still have no certain grounds for pride here, however; for what is learned in schools and, consequently, what is being assessed here by NAEP may be so irrelevant to a citizen's life purposes and activities that he has no reason to learn any more of it. Also what is being noted here as learning may be a consequence of experiences at a certain age, but not necessarily school experiences.

Another tragedy is depicted by Exhibits 3, 4 and 5. The gross differences and inequalities in our society that attend race and class, for the most part, are part and parcel of schooling, learning and testing. That, of course, is hardly a startling revelation. Everybody knows that the racist and classist biases in this society are pervasive. They operate in schools quite as easily as anywhere, despite 20 years of desegregation experience. However, buried in the tangle of lines in Exhibit 4 is the whisper of a hypothesis that, assuming a part of the difference between 13 and 17 is a schooling effect, blacks and the poor may get something of a break in schools. While the blacks, the low metros (these are mainly the inner city poor and probably are largely black) and those whose parents never attended high school have the three lowest median raw p-values of all 13-year-old groups (around 60%) in the Rights and Duties color scheme, they show the most precipitous gains by age 17. The low metro groups show a 20 percentage point gain between age 13 and age 17; the black samples and the no high school samples show 13- and 15-point differences. The national difference is only 4 points and no other group gained more than 8.3%. The gains of the low metro, no high school and black groups may be more of a function of starting points. It is much easier to post gains with a starting point of 59% than it is with a starting point of 77%. Their gains are short lived, however; in the adult world, the black and low metro categories, along with the Southeast region, show group disadvantages of from 6 to 12 percentage points below the national median.

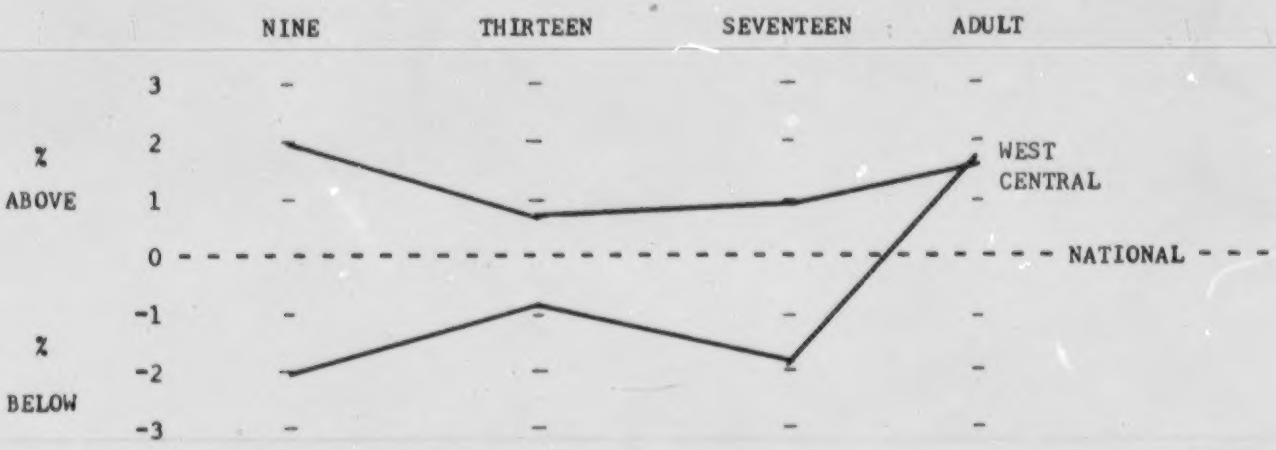
An alternative hypothesis may be more powerful. These 17-year-old blacks in 1971-72 had been witness to and had probably participated in a dramatic civil rights revolution in our society throughout their school life. They, more than most, would have been sensitized to such issues since 1954, the year in which some of them were born. In this hypothesis, the schools can claim no credit, for it was these young 17-year-olds who brought the pertinent concepts to the school; they didn't learn them there.

There is also the possibility that the gains for black, no high school and low metro groups reflect a condition noted in other studies — e.g., the Columbia University Citizenship Education Project — that black youths and some others learn more about politics and the like in school because they have less opportunity for formal learning about institutions outside of school than do white, middle-class youths and some others.

## Comparison of Group Differences from National Success Levels in Social Studies and Citizenship Assessments

In the following exhibits and discussions, group performances will be reported as delta p-values for individual exercises and median delta p-values for groups of exercises or color schemes. When delta p-values are used, the performance of a selected group is shown as deviating by so many percentage points above or below the national raw p-value on a given exercise. In the exhibits in this section, we will adopt a NAEP convention by showing the national performance levels as lying along a horizontal line designated as 0. It is not useful to identify national raw p-values in the exhibits since the median delta p-values charted in the figures for two or more groups at any age level most likely report scores on different exercises. Exhibits simply display the typical performances of groups. The performance level of the groups involved will be charted as broken lines that connect the median delta p-values for each age level above and below the 0 national line. Exhibit 6 is an example of this convention.

EXHIBIT 6. Knowledge of Institutions - Regions - Median Delta P



In Exhibit 6, the Central region 9-year-olds are 2 percentage points above the national raw p on their median exercise while Western 9-year-olds are 2.1 percentage points below the national level on their median exercise, etc. In our subsequent discussions, we may, for convenience and relief, use shortened versions of such statements, e.g., Central 9-year-olds are 2% above the national percentage while Western 9-year-olds are -2.1% below.

Our first set of exhibits will depict the performances of selected groups on the total number of exercises included in this chapter. The references will be to "Knowledge of Institutions," which is treated here as a color scheme of related exercises. In four instances it has been possible to make some general comparisons with the NAEP's citizenship assessment made in 1969-70. Goal D of the citizenship assessment objectives, "Knowledge of Governmental Structures and Functions," is similar to the overall color scheme, "Knowledge of Institutions," used as the basis of this chapter. There are several difficulties with the comparison, not the least of which is that the citizenship assessment results are reported in such enigmatic ways that many values simply cannot be found and all are rounded estimates for group results. The major difficulty, of course, is the doubtful fit of the two sets of exercises.

### *Regions*

Exhibit 7 depicts the performances of the four age levels on all exercises in the regional categories. The Exhibit 7 display shows very little variance among the four regions, though on most of the exercises, the Southeastern adults are significantly below the national level. This is the most dramatic difference in the exhibit, though the difference between Northeastern and Southeastern 13-year-olds is exactly the same as the difference between the adults in the two regions.

The real puzzle in Exhibit 7 lies with the Western 17-year-olds and adults. At 17, all but 5 of the 42 exercises included here are below the national percentages; one fourth are significantly below. Among the adults, 80% of the exercises are above the national percentages, and one fourth are significantly above. Perhaps the high-scoring Northeastern teenagers of a decade ago moved West. That hypothesis is of dubious quality, obviously; at least, the pattern does not show up in the citizenship assessment of 1969-70, which is displayed for regions in Exhibit 8. The same general, fairly flat configuration of Exhibit 7 is apparent also in Exhibit 8 with the Northeast and Central lines lying wholly above the national percentages and with the Southeast lying wholly below. The dramatic rise in the West from -0.5 at age 13 to 1.7 at age 17 is not altogether different than what occurs between age 17 and adult in Exhibit 7.

### *Race*

Exhibit 9 shows the performances of the four age samples on all exercises as categorized by black and white races. Excluded from these displays are about 125 to 220 persons in each of the samples of some 2,500 persons who were classified as "other." The "other" category includes Orientals, who qualify as racially distinct, and some ethnic groups who do not. The category is too indistinct to characterize accurately and too small to include in this breakdown. In National Assessment's testing procedures, racial membership is determined visually by the test administrators in the field. That is part of the reason for the confused "other" category. All field administrators do not share the same definitions of race. Also included here in Exhibit 10 are the group results for blacks for Goal D in the 1969-70 citizenship assessment. Neither national percentages nor white median delta p-values are reported for the citizenship assessment. Nonetheless, the results are obviously tragically similar.

As shown in Exhibit 9, the success difference between the whites' and blacks' median delta p-values at every age level is from 13% to over 20%. The exercise story is the same at

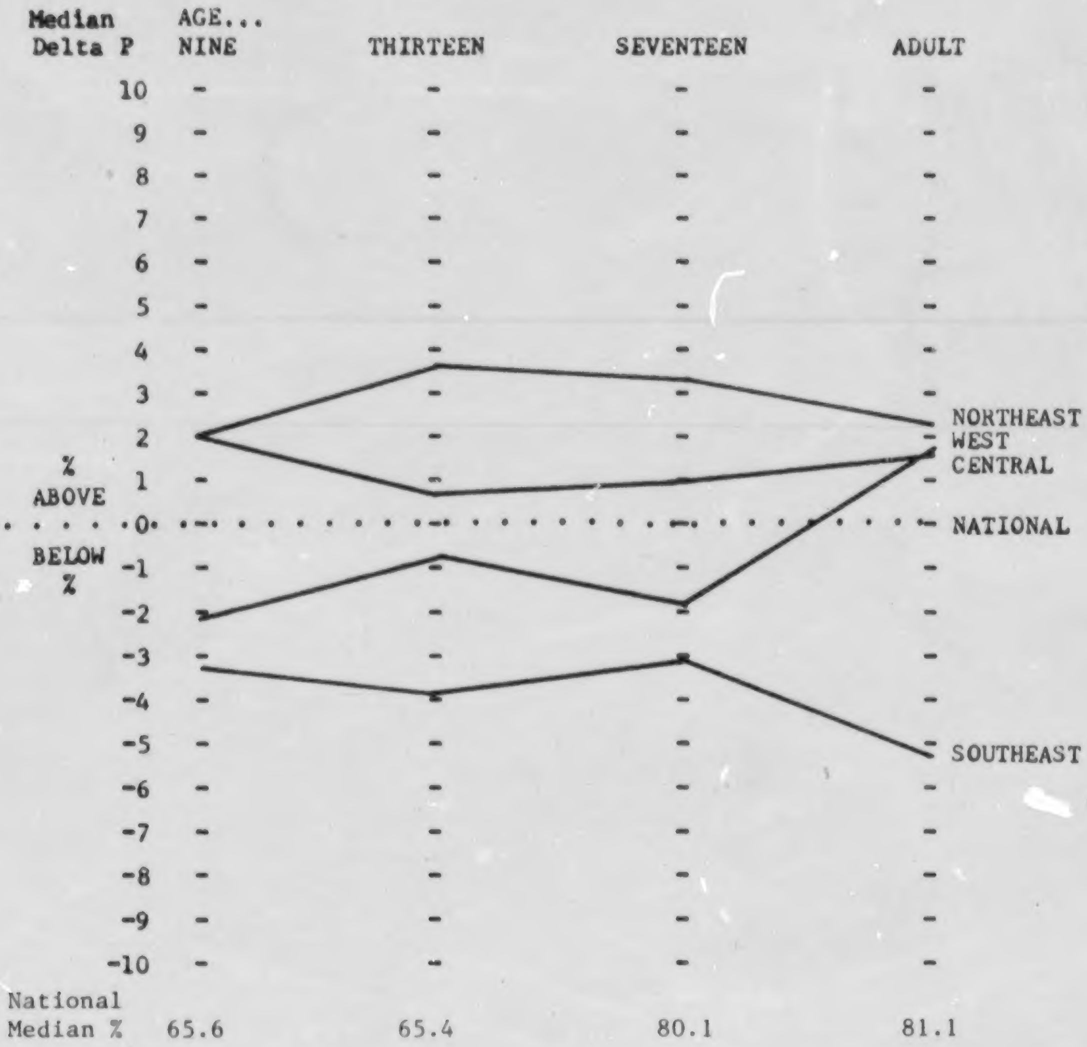


EXHIBIT 7. Knowledge of Institutions - Regions - Social Studies 1971-72

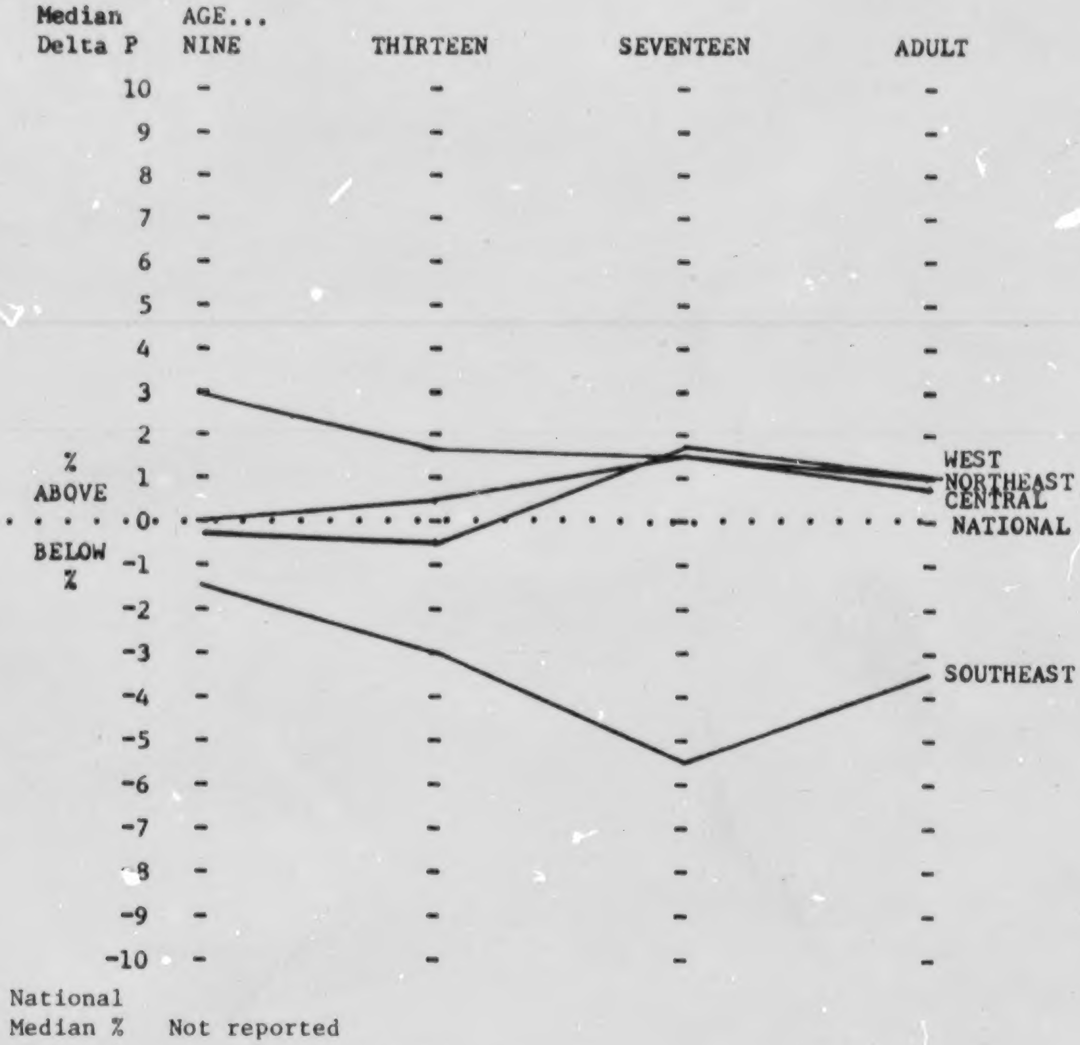


EXHIBIT 8. Knowledge of Governmental Structures and Functions - Regions - Citizenship 1969-70

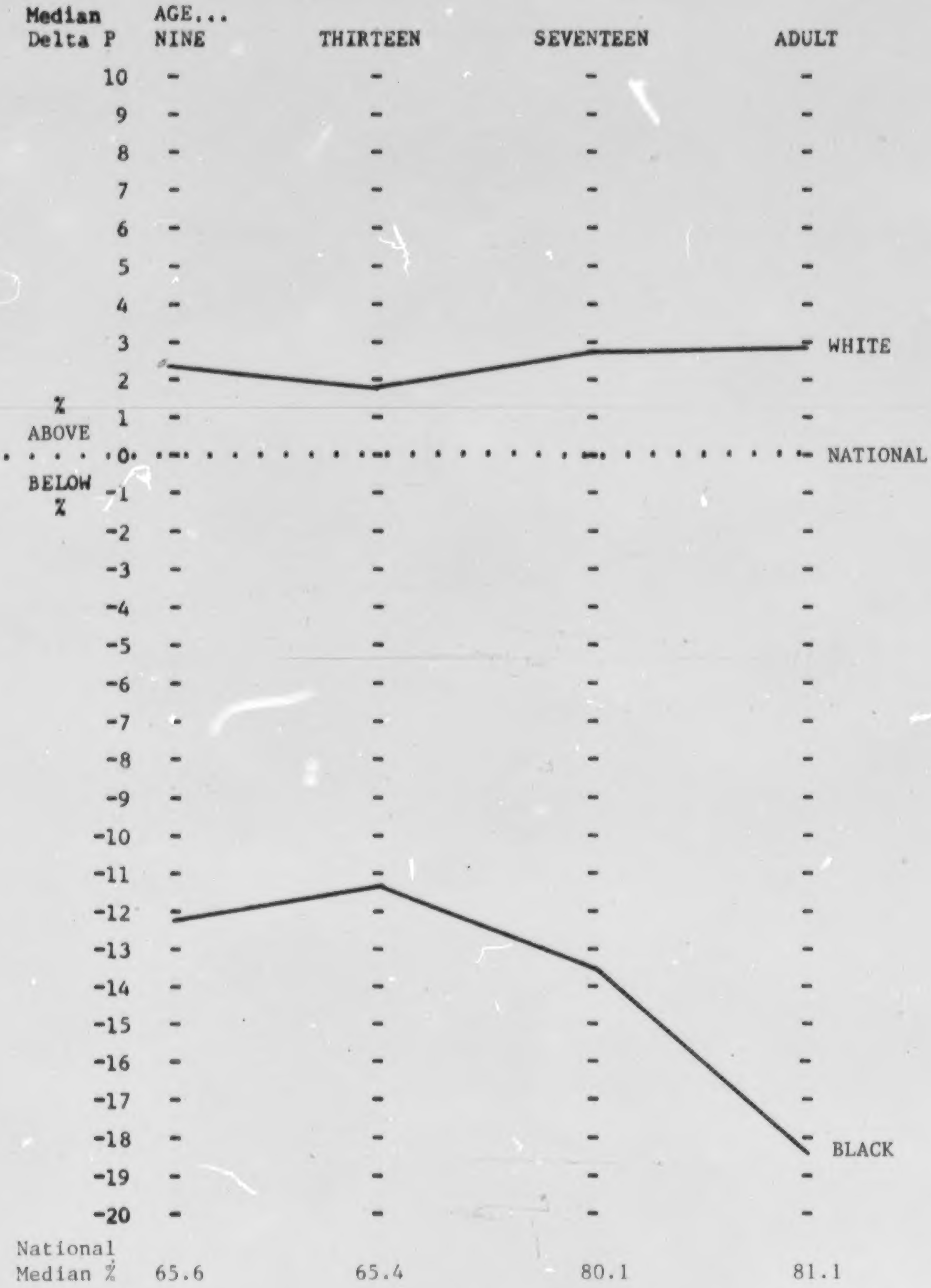
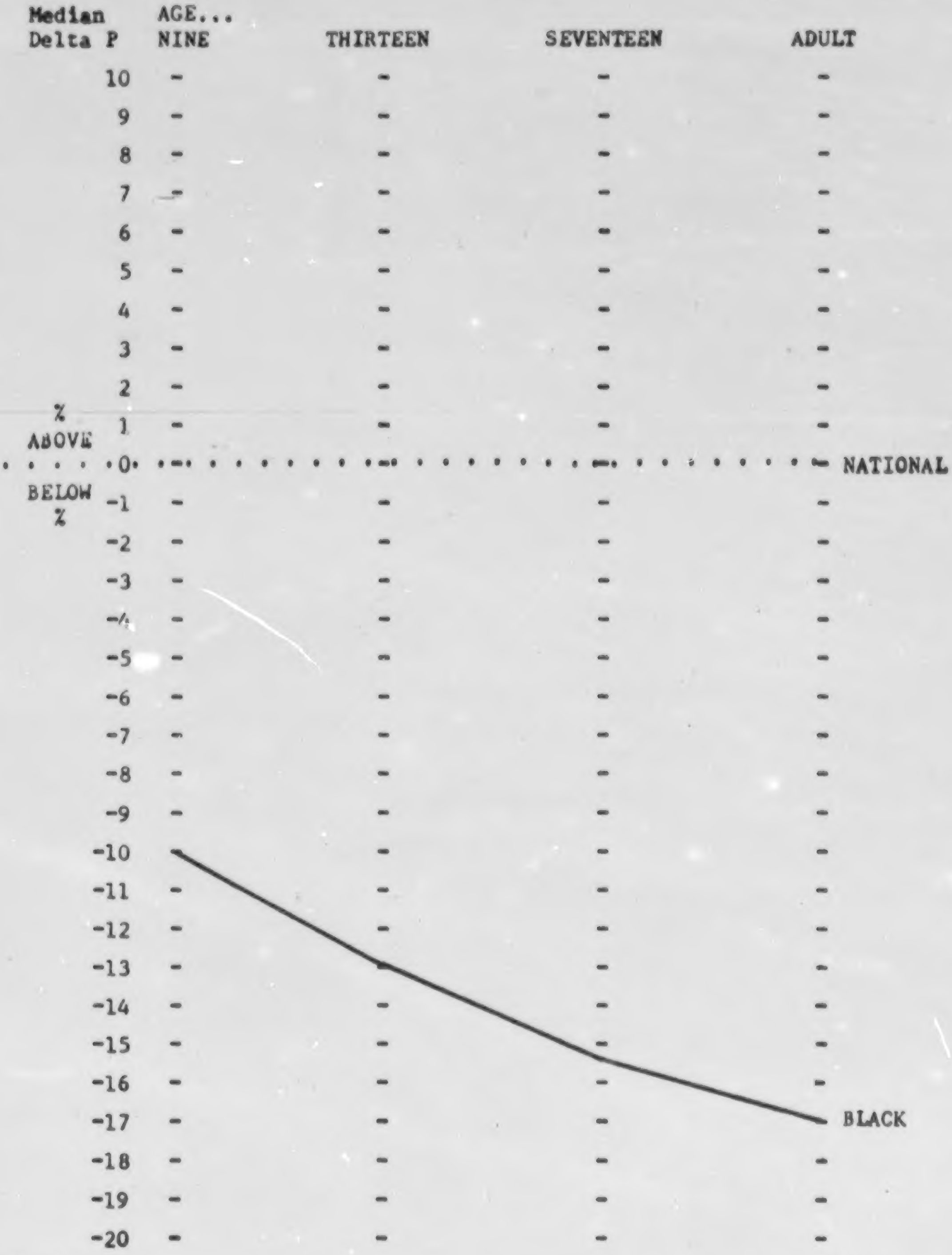


EXHIBIT 9. Knowledge of Institutions - Race - Social Studies 1971-72



National
Median %   Not reported

EXHIBIT 10. Knowledge of Governmental Structures and Functions - Race - Citizenship 1969-70

every age level. At age 9, the whites have 6 out of 6 exercises significantly above the national percentages; the blacks have 6 out of 6 significantly below. At age 13, the whites have 24 out of 26 exercises above the national percentages, 18 of them significantly above. The blacks, on the other hand, have 23 out of 26 exercises below the national percentages, with 20 significantly below. White 17-year-olds have 39 out of 42 exercises above the national percentages while the black 17-year-olds have 40 out of 42 exercises below. Of these, 37 are significantly above for whites and 36 are significantly below for blacks. Adult whites have 39 out of 41 above and adult blacks have 39 out of 41 below the national percentages. All but one delta p above and two below are statistically significant. On several individual exercises at the adult level and on a few at other levels, the difference between black and white raw p-values is more than 30%. In one case the difference is nearly 40%; a white adult, in that case, was two and one half times more likely to answer correctly than a black adult. As measured by this social studies assessment, the pattern emerges of two racial groups in this country dramatically and tragically diverse in their knowledge and understanding of the basic legal and political institutions that order and protect their lives and interactions.

### *Parental Education*

Exhibits 11 and 12 display breakdowns of the social studies assessment results on Institutions and the earlier citizenship assessment results on Governmental Structures and Functions (Goal D) by parental education. A respondent was placed in the post high school category if one or both of his parents attended some school beyond high school; in the high school graduate category if one or both of his parents graduated from high school; in the some high school category if one or both of his parents attended high school, but neither graduated; and in the no high school category if neither parent went to school beyond the eighth grade.

It is well known that father's education has important social class membership effects in this society, so that kind of differentiation is obviously working in these arrays. A distinctive characteristic shared by the two data sources is the absolute and hierarchical discreteness of the categories across the age levels. The lower-education groups never close the gap. The some high school adults show a precipitous gain in Exhibit 12, however, which would be interesting to explore. It could reflect nothing more than a chance sampling fluctuation; but in the midst of the parallelism otherwise displayed in this exhibit, the results of this category across the age levels are anomolous. The parallel upswings of all adult categories in Exhibit 11 will be looked at more specifically in the analyses in the next section of this chapter.

### *Size and Type of Community*

Exhibits 13 and 14 are also related in that they show the two sets of data categorized by size and type of community. These categories also have heavy socioeconomic effects in that high metro is mainly an affluent grouping; urban fringe is suburbanite, presumably middle-class America; extreme rural is primarily a farm population; and low metro is inner city, thus mainly, though not exclusively, city Black.

It is tempting to overwork these data and to extend unreasonable assumptions of reliability and validity. But without invoking validity, it seems warranted to say that the exercises used in the 1969-70 citizenship assessment and those used in the 1971-72 social studies assessment discriminate among these STOC groups in the eight age samples involved in very similar ways. The four community types selected here actually include only about 45% of the data; but for the social studies assessment, the remaining results for main big city, medium city and small place are all within 1.5% of the national percentages; most are less than 0.5%. In the citizenship assessment, all but two of the results not shown on Exhibit 13 are less than 2.5%. The high variances are included in these displays. They are unquestionably sensitive to socioeconomic influences.

Though in the next sections, where the two social studies color schemes will be treated separately, we will treat the validity question more explicitly, it is transparent here, if we assume test validity, that the social classes in this country vary greatly in their knowledge and understanding of the institutions they live by. Before this classist hypothesis is embraced, however, one should attend carefully to the judgments reported by both Larkins and Hunkins concerning the validity of objectives and exercises comprising the NAEP social studies and citizenship assessments. Larkins, in particular, suggests the presence of social-class effects in these assessments, but they are at least as evident in the assessors as in the assessed. Schooling, we might infer from the squeezing of the variance that occurs at age 13 and age 17, has some effect toward leveling; but once the influence of the school is gone, the influences of the wider milieu surface. Most of the high metro adults will have gone to college; most of the low metro adults will not. That in itself could account for some difference. Either the exercises are so profoundly culturally biased that the lower classes have a built-in content and language disadvantage while the upper classes have built-in advantages; or the reading, conversation and viewing habits, along with other traits, of the high and low metro groups are so diverse that matters attended to in one are more in line with what these exercises refer to while the matters attended to in the other are outside this realm.

But these two explanations are the same. They both say that apparently the content of these questions is reinforced by the experiences of one group after school and not reinforced by the experiences of the other. From this point on the consideration becomes normative. One may say with Robert Coles that the effects of the inner city are pathological and possibly with James Coleman that one set of values is more appropriate for life in this country; or one may say that neither National Assessment nor any other white middle-class organization — schools, for example — can observe, teach or assess the lower class in this country with accuracy and fairness.

In the following section we will analyze the social studies data by the two smaller color schemes with more reference to individual exercises.

## Analysis of the Social Studies Data by Knowledge of Structures and by Knowledge of Rights and Duties

Results reflecting the performances of the four age level samples and analyzed by region, sex, race, parental education and size and type of community for the two color schemes, Knowledge of Structures and Knowledge of Rights and Duties, are displayed in Exhibits 15 through 24. The

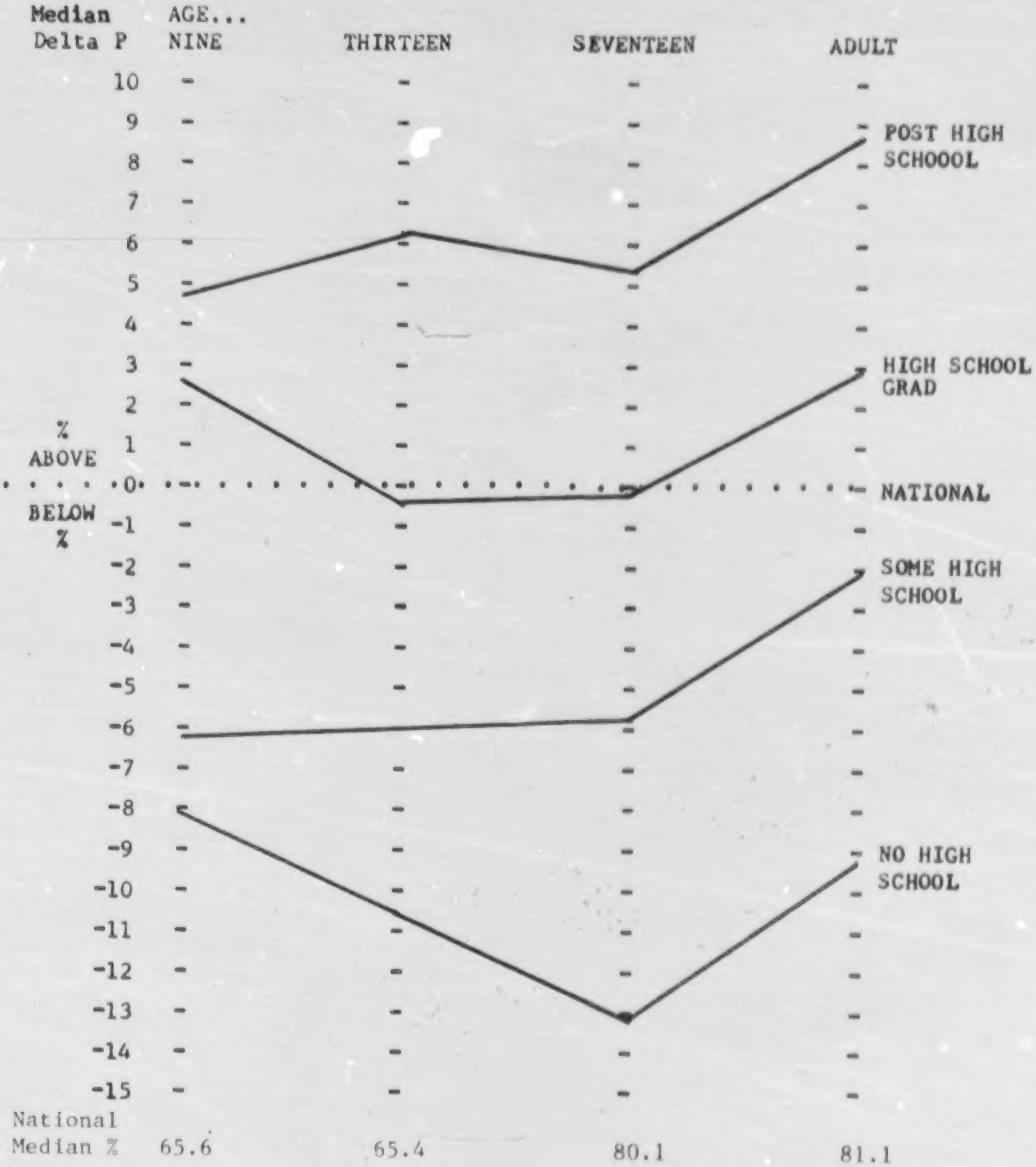


EXHIBIT 11. Knowledge of Institutions - Parental Education - Social Studies 1971-72

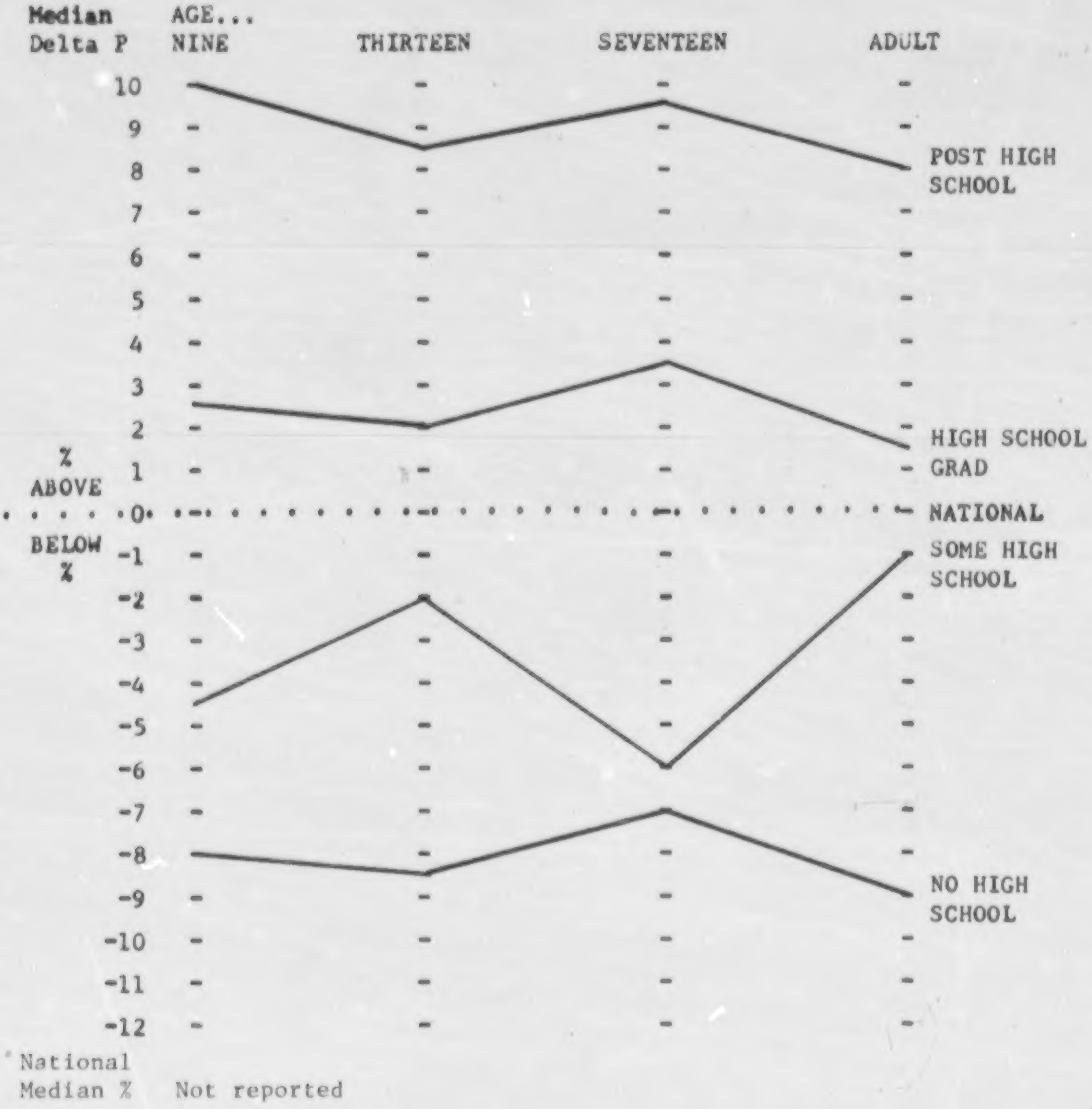


EXHIBIT 12. Knowledge of Governmental Structures and Functions - Parental Education - Citizenship 1969-70

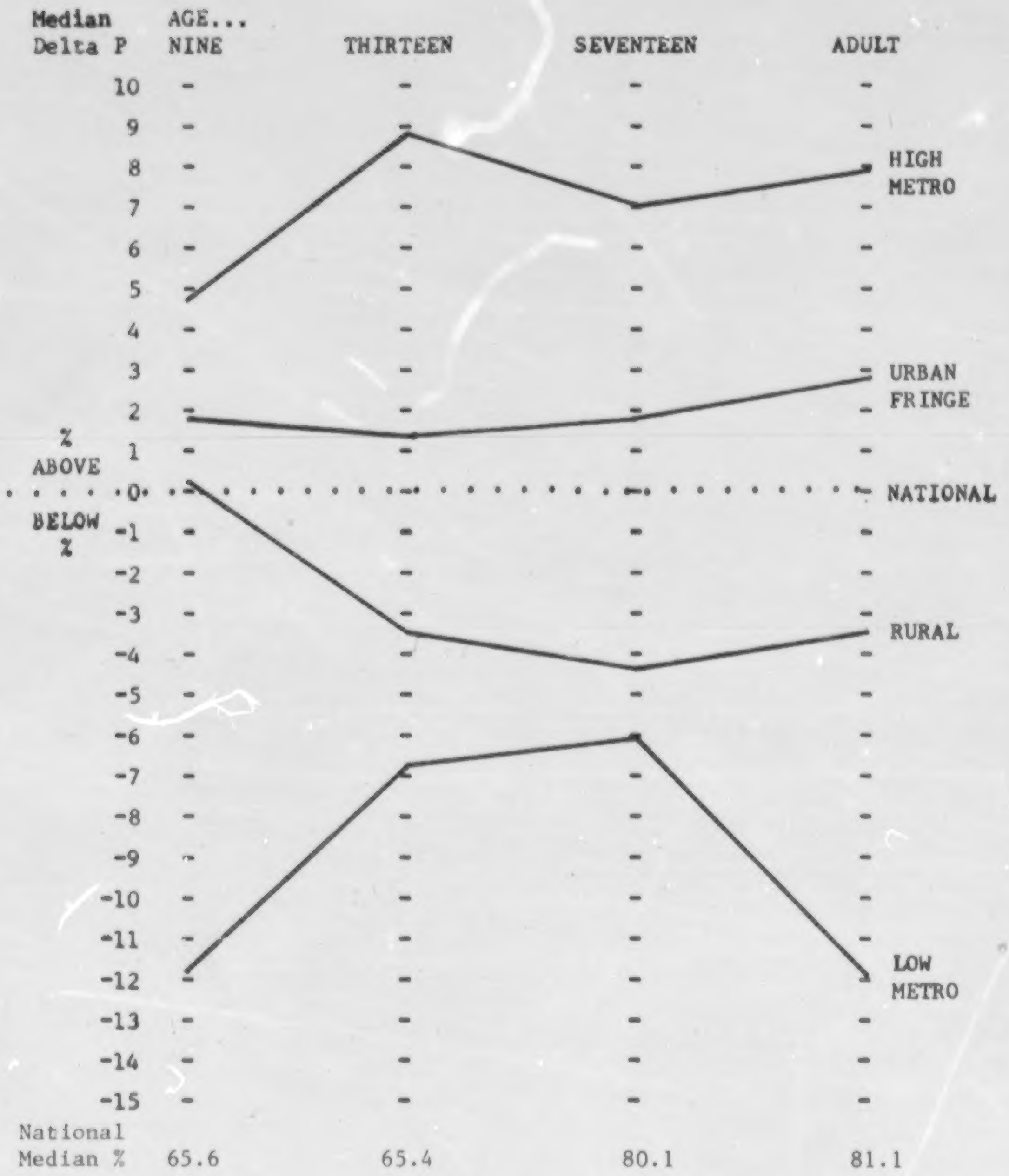
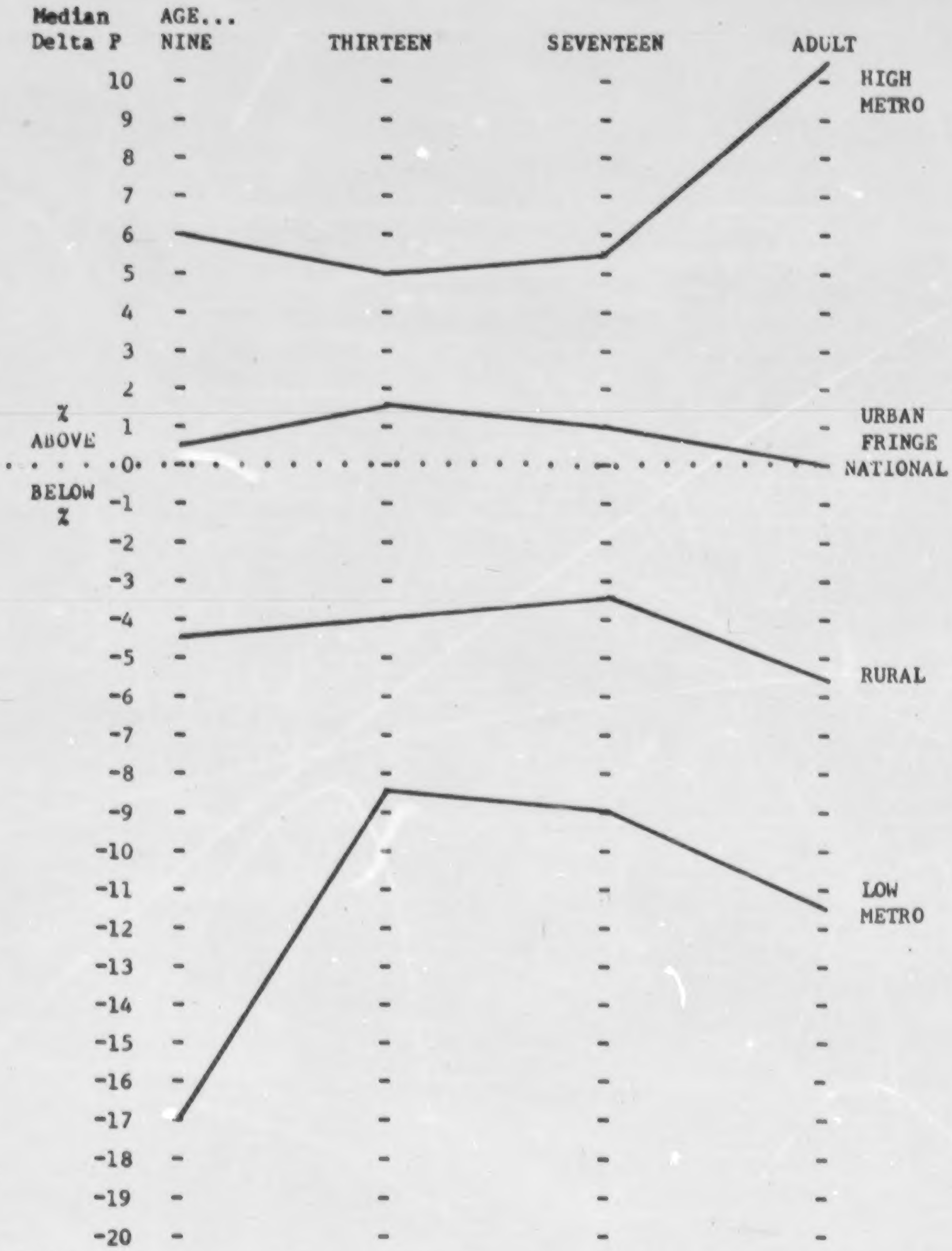


EXHIBIT 13. Knowledge of Institutions - STOC - Social Studies 1971-72



National
Median %   Not reported

EXHIBIT 14. Knowledge of Governmental Structures
and Functions - STOC - Citizenship 1969-70

same statistical treatment is used, i.e., national percentages are shown in exhibits as 0 and group effects on color schemes are shown as median delta p-values; however, each color scheme is based on only a portion of the total number of exercises. The presumption is that each of these color schemes is more precisely homogeneous than the total group of exercises. That, of course, is a visual judgment and may be checked in Appendix C where the texts of the released exercises and the topics of the unreleased exercises are shown. The results in this section are based on the following number of exercises:

| | Age 9 | Age 13 | Age 17 | Adult |
|---|---|---|---|---|
| **Structures** | 5 | 16 | 21 | 21 |
| **Rights and Duties** | —— | 10 | 21 | 20 |

## *Regions*

The first set of exhibits in this section is made up of Exhibit 15, which displays the Structure color scheme, and Exhibit 16, which displays the Rights and Duties color scheme for each of the regional categories of Southeast, West, Central and Northeast.

*Structures.* As Exhibit 15 clearly shows, the variance across regions for the exercises assessing structural knowledge is small. The median delta p-values are all less than 5 percentage points. Most are less than 2%.

At the 9-year-old level, the exercise showing the biggest difference between Northeast and Southeast youths asked,

> Which one of the following is usually the head of a government in a town?
> The mayor
> The governor
> The chief of police
> The school principal
> I don't know.

Nationally, 57.5% of 9-year-olds know the correct answer; 48.1% of the 9-year-olds know it in the Southeast, and 66.7% of the Northeast 9-year-olds know it. Both values are significantly different than the national percentage and both are atypical responses for the regional groups. A simple cultural effect may be operating in the question due to the different geographic profiles of the two regions, the different names given to town heads — e.g., first selectman and town board chairman — and even different names for towns in some states — e.g., villages.

At age 13, the West and Central are more alike. As populations they are indistinguishable, in fact, with only two atypical responses and three significantly different than the national percentage between them. The Northeast and Southeast are more different from each other, though the contrast is far from dramatic. On most exercises the two groups of 13-year-olds are between 5 and 10 percentages points apart with the Southeast in all but one case being below the national percentage and the Northeast in all but two cases being above. They are within 3 percentage points of each other on a question about financing the government. The text for this question is unreleased. (To provide a basis for future comparison, NAEP holds in secret about half the exercises used in an assessment. We make reference to these unreleased exercises by topic only.) At any rate, fewer than half of all 13-year-olds know the answer (national percentage — 47.2). Northeastern youths have a delta p of -1.6%; Southeastern 13-year-olds, -4.58%. The two groups are farthest from each other on an unreleased exercise on school governance. Northeast 13-year-olds are 6.2% above the national percentage while Southeast 13-year-olds are -6.3% below.

At age 17, the two most contrasting groups are again the Southeast and Northeast, though the Western youths scored the lowest of all on two unreleased constitutional questions to provide the greatest contrast for the Northeast with differences of 16 and 19 percentage points. In another unreleased question on foreign affairs the Southeast performed the best, at 5% above the national percentage, while the West again is lowest, at -6.1% below the national percentage.

At the adult level, the Northeast, Central and West are indistinguishable with respect to median delta p-values. The Southeast, whose adults make the poorest showing here, is below the national percentages on all 21 exercises by only -2% to -8%. The other regions have only four or five results below the national percentages. None, however, is more than 6.8% above.

The range of performance for all adults in this color scheme is from a poor national percentage of success of 45.7% on an unreleased exercise on a constitutional provision to a whopping 96.3% on another unreleased question on the armed forces. About two thirds of the entire adult sample was able to answer a half dozen unreleased questions on federal government details. A released question, typical of these detail questions on the federal government, asks,

> Which one of the following has the power to declare an act of Congress unconstitutional?
> The Congress
> The President
> The United States Supreme Court
> The United States Department of Justice
> I don't know.

Nationally, 61.8% of the adults answered correctly. Delta p-values range from 3.0% for Central adults to -4.5% for Southeastern adults.

*Rights and Duties.* Exhibit 16 displays the median delta p-values for the four regional categories on the color scheme, Knowledge of Institutional Rights and Duties, for age levels 13, 17 and adult. No 9-year-old results are referred to in this color scheme because only one exercise was classified under Rights and Duties.

Released exercises in this color scheme involve such matters as atheists holding office, criticism of officials and the military by newspapers and others, picketing rock concerts and police, assembly in parks, congressmen seeking views of constituents, racial discrimination in employment and laws against vandalism. Unreleased exercises relate to freedom of press and religion, property and petition rights, and government criticism. The color scheme as a whole is presumed to inquire into knowledge and understanding of individual rights and duties in several institutional arrangements, mostly political, in this society.

Most of the exercises in this color scheme may be interpreted as assessing a respondent's commitment to some democratic values. A released example of this sort of question asks,

> Should a person who does not believe in God be allowed to hold a public office?
> Yes
> No
> Undecided.

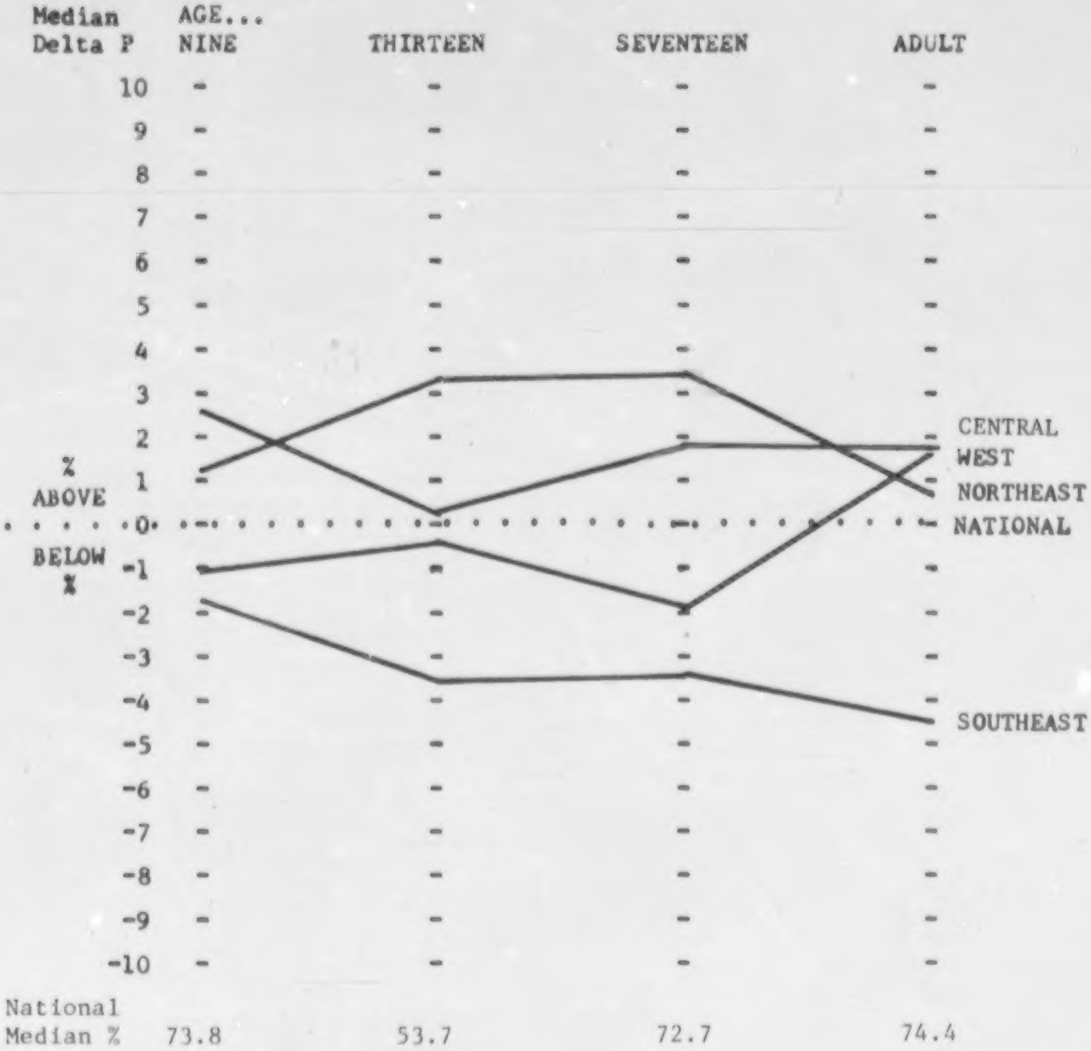


EXHIBIT 15. Knowledge of Institutional Structures - Regions - Social Studies 1971-72

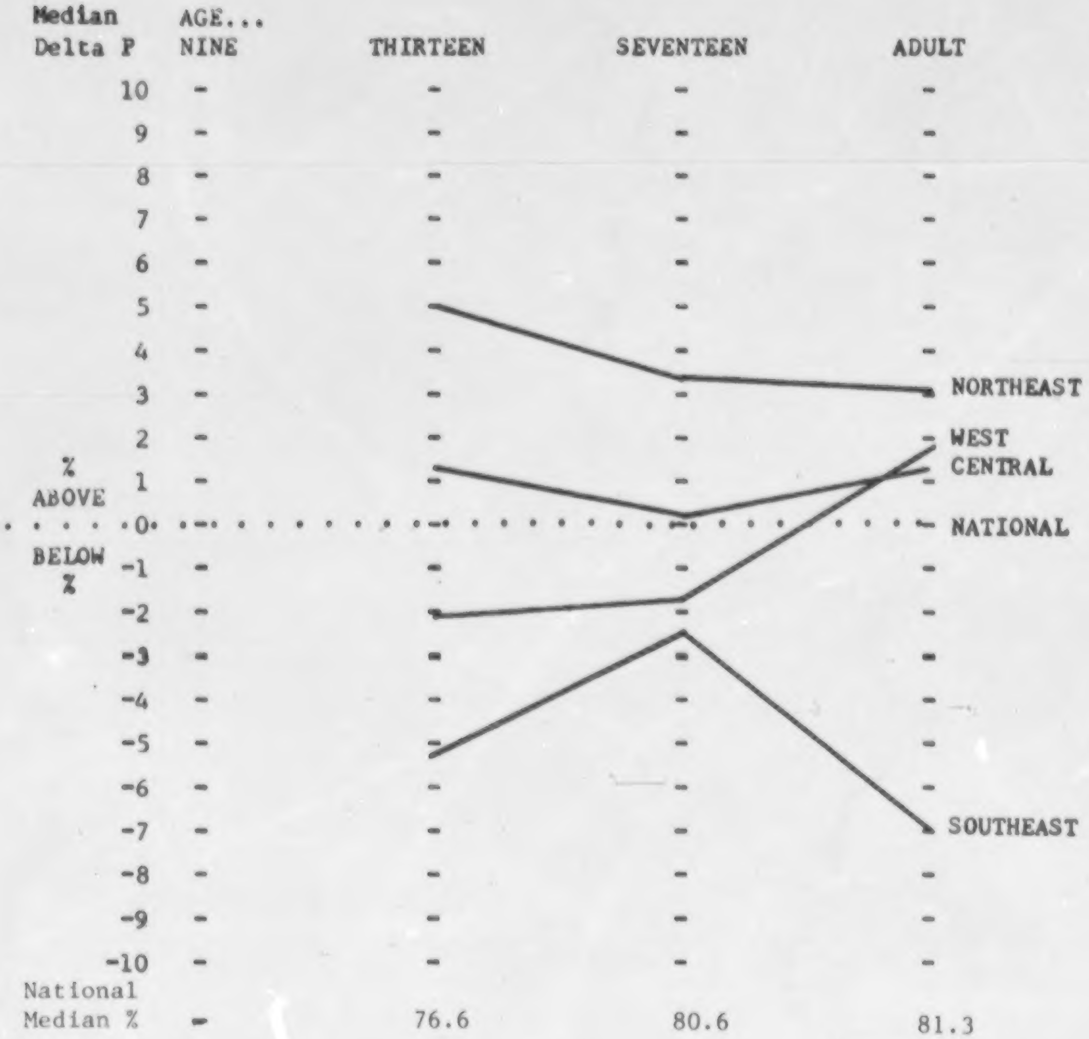


EXHIBIT 16. Knowledge of Institutional Rights and Duties - Regions - Social Studies 1971-72

From that interpretation, it seems perfectly appropriate to think of the set of questions as inquiring into the democratic beliefs of the persons being questioned. Our own reluctance to call the color scheme by such names as attitudes, beliefs and commitments is based on two reasons. First the exercises refer, for the most part, to rights and duties that are fully legal and established in this country. They do not present issues that are presently the subject of hot debate in the society, nor do they refer to any controversial extensions of democracy into new areas. Second, it is very difficult to ascertain the actual realm of assessment with such questions. The recent literature on value analysis only serves to emphasize the enigmas involved. Without further questioning and exploration, there is no certainty whether the question elicits descriptive or valuative information.

To the credit of NAEP, additional information has been requested on some of these *should* questions to help insure their operation in the value realm. The respondents are asked to express reasons for their answers. The following is a released example.

> A. Do you think people should be allowed to picket the holding of a rock festival as a protest against it?
>
> Yes (go to B)
> No (go to B)
> Undecided (go to B)
> No response (after 10 seconds, go to C)
>
> B. Please give a reason for your answer.

Again, all we can say with respect to this strategy is that it is perhaps as good as can be done with written exercises. The further assumption is required, however, that the respondent's own criterion is being expressed when he offers a reason. Otherwise, we are back to square one and still uncertain whether descriptive or normative information is being expressed. By design, we excluded the reason giving parts of such questions from this *knowledge* color scheme in anticipation of having another color scheme in which they would be featured. This value color scheme, as we stated earlier, was among those not computerized. Its absence, we recognize, places this Knowledge of Institutional Rights and Duties color scheme, as well as this interpreter, in a vulnerable condition.

As Exhibit 16 clearly shows, the regions consistently maintain their respective positions in the area of Rights and Duties. That suggests that the two color schemes, Structures and Rights and Duties, are not discrete, i.e., they are probably measuring similar things. Nonetheless, the regional groups are more spread out in this color scheme. Over 10 percentage points separate Northeast and Southeast median delta p-values at both age 13 and adult levels. The Central and Western categories at each age level continue to look very much alike and continue to be the norming group. Only three results among all delta p-values for both groups for all ages are more than 5% above or below the national percentages.

A selected set of exercises (Table 9) will illustrate how the Northeast and Southeast differ. The released questions are represented here by shortened texts. Since the values reported in these examples are delta p-values and national percentages for individual exercises, rather than median values as in the previous displays, they are directly and specifically comparable. The Northeast and Southeast scores can be added to and subtracted from the national percentages.

## TABLE 9. Northeast/Southeast/National Comparisons on Selected Exercises

| | Age 13 | Age 17 | Adult |
|---|---|---|---|
| **Should an atheist hold public office?** | | | |
| Northeast | 11.6 | 14.2 | 7.4 |
| Southeast | -16.8 | -16.1 | -20.3 |
| National % | 59.4 | 62.6 | 55.6 |
| **Should an atheist be allowed to express his views publicly?** | | | |
| Northeast | 7.6 | 2.3 | 3.0 |
| Southeast | -8.4 | -4.9 | -10.1 |
| National % | 63.9 | 78.1 | 77.9 |
| **Should newspapers criticize public officials?** | | | |
| Northeast | 8.4 | 7.8 | 2.9 |
| Southeast | -8.9 | -9.3 | -7.0 |
| National % | 48.9 | 73.4 | 81.1 |
| **Should people be allowed to picket against a rock festival?** | | | |
| Northeast | | 6.9 | 0.6 |
| Southeast | | -9.3 | -8.6 |
| National % | | 52.7 | 63.0 |
| **Should people be allowed to assemble in parks to make demands?** | | | |
| Northeast | 5.6 | 3.4 | 5.0 |
| Southeast | -11.7 | -4.6 | -7.8 |
| National % | 63.5 | 87.2 | 77.5 |

The selected set of exercises above cannot be used appropriately to characterize the groups. Exhibit 16 does that more reliably. However, some generalizations and contrasts inevitably emerge.

In a country that is presumably built on democratic principles, some of those believed by many to be fairly basic fail in these exercises to receive anything like unanimous approval. If one prefers an alternative interpretation, a great many persons who live in this country do not know what its basic principles are. From another view, however, one can say that 75-to-80% support for anything in a country as pluralistic as this one is nothing short of miraculous. Moreover, the application of a value conflict model to these data would likely contribute to their understanding.

Whatever generalization is preferred, it would appear fairly certain that a person who professes not to believe in God could not be elected dog catcher of Orange County, North Carolina, on a bipartisan ticket. Atheist politicians would not seem to be shoo-ins anywhere in the country. That exercise result, along with one about allowing persons to picket a police station to protest police brutality and another about knowing that prayer is excluded from schools on the grounds of church and state separation, barely made a majority among adults. Picketing and the Supreme Court's religion-banning decision fared no better among the nation's 17-year-olds. Fewer than half chose the separation principle for the one and a bare majority would allow pickets to protest either police brutality or a rock festival. In the latter cases, rejection of the time honored "right" of protest over that range of social offenses must mean something.

The three age levels were asked,

Should race be a factor in hiring someone for a job?
Yes
No
I don't know.

On this very important question, we take heart. Over 90% of every category at every age level, except Southeast and West 13-year-olds and Southeastern adults know that racial discrimination in employment is illegal or believe that race should not be a factor in employment, whichever interpretation one prefers. A difference between the Southeastern and Western 13-year-olds is that for the Southeasterners this result is next to their best performance relative to the national percentages. They are atypically high. But for the Western 13-year-olds, this result is their poorest performance on all Rights and Duties exercises. The delta p-value of -6.4% is significantly below the national percentage of 88.8% for all 13-year-olds.

A visual comparison of the rank ordering of exercises according to delta p-values, a contribution of the computer that will not be shared in this report, for each of the age levels and region groups suggests, much like Exhibit 16, that Southeast and West youth are much alike; Northeast and Central youth have much in common; and Northeast, Central and West adults are very similar. The dramatic events that occur in this breakdown of the data across regions are the accommodation of eastern liberalism by West adults and the retrenchment of Southeast adults in a conservative, Bible-belt, law-and-order posture.

## Sex

Exhibits 17 and 18 display the results in the Structure and Rights and Duties color schemes according to sex. Our first examination will concentrate on the Structure questions.

*Structures.* Exhibit 17 shows the median delta p-values for males and females across all ages for the Structure color scheme. According to the median delta p-values used to characterize males and females at ages 9 and 13, the two groups of young persons are not very different in their knowledge of the structures of institutions. In comparing the two groups at ages 9 and 13 on individual exercises, the similarity judgment is only partially upheld in that the two groups are significantly different than the national percentage on 7 of 16 exercises at age 13 and 2 of 5 exercises at age 9; but in none of these is the difference more than 7 percentage points. So at ages 9 and 13, boys and girls are different in what they know, believe or hold important; but they are not dramatically different. Differences this small may in fact be only chance products.

Specifically, at age 9, males are 5% more likely than females to know that mayors and not police chiefs or school superintendents are the heads of towns. Also boys are about 5% more likely than girls to know that the health department is involved in restaurant inspection, not selling food or putting out fires. The importance of these pieces of information seems less important than the facts that only 57% of all 9-year-olds know about town mayors and only 35% know about restaurant inspection. Even these results seem explainable within the normal world of 9-year-olds.

By age 13, males and females seem quite similar as shown in Exhibit 17. In fact, however, at least on some items, they are more different than at age 9. The males outpoint the females from 6 to 7 percentage points on questions relating to national and international governmental structures, e.g., what the United Nations does and how a presidential candidate is nominated in the United States. But even with their superior performance in the latter case, only one out of five of the boys knows about national political conventions.

By age 17, a pattern that looked possible at age 13 seems more established. The depoliticization of females is in full progress by age 17. While the differences between males and females on individual questions are not great — about half the significant differences are between 5 and 10 percentage points — they are significantly different on 14 exercises. On 13 of these the males are above the national percentages while the females are below. The greatest male advantages are on exercises referring to such matters as the nomination of presidential candidates and the declaring of congressional acts unconstitutional. The only significant female advantage is posted in a set of exercises requiring the interpretation of a replica ballot, as follows (national percentages of success and female and male delta p-values are indicated in parentheses).

"The ballot below was used in a general election." Look at the ballot to answer the questions on this and the following two pages.

| | | | | | |
|---|---|---|---|---|---|
| | LEGISLATIVE | | COUNTY | | |
| OFFICES | SENATOR IN CONGRESS (vote for one) | REPRESENTATIVE IN CONGRESS (vote for one) | COUNCILMAN (vote for two) | | TAX ASSESSOR (vote for one) |
| DEMOCRATIC | Alan F KIRK | John G SMITH | Martha G DAVIS | Peter V MOSS | |
| REPUBLICAN | James M JONES | Mary O'CONNOR | John RICHARDS | Michael M MERWIN | Joseph L LASKI |

A. If you wanted to vote for Kirk for senator, could you also vote for O'Connor for member of the House of Representatives?

(National % 72.8)
(Female -1.34; Male 1.41)

[✓ Yes]
No
I don't know

B. Could you vote for both Davis and Moss for councilman?

(National % 83.2)
(Female 2.11; Male -2.23)

[✓ Yes]
No
I don't know

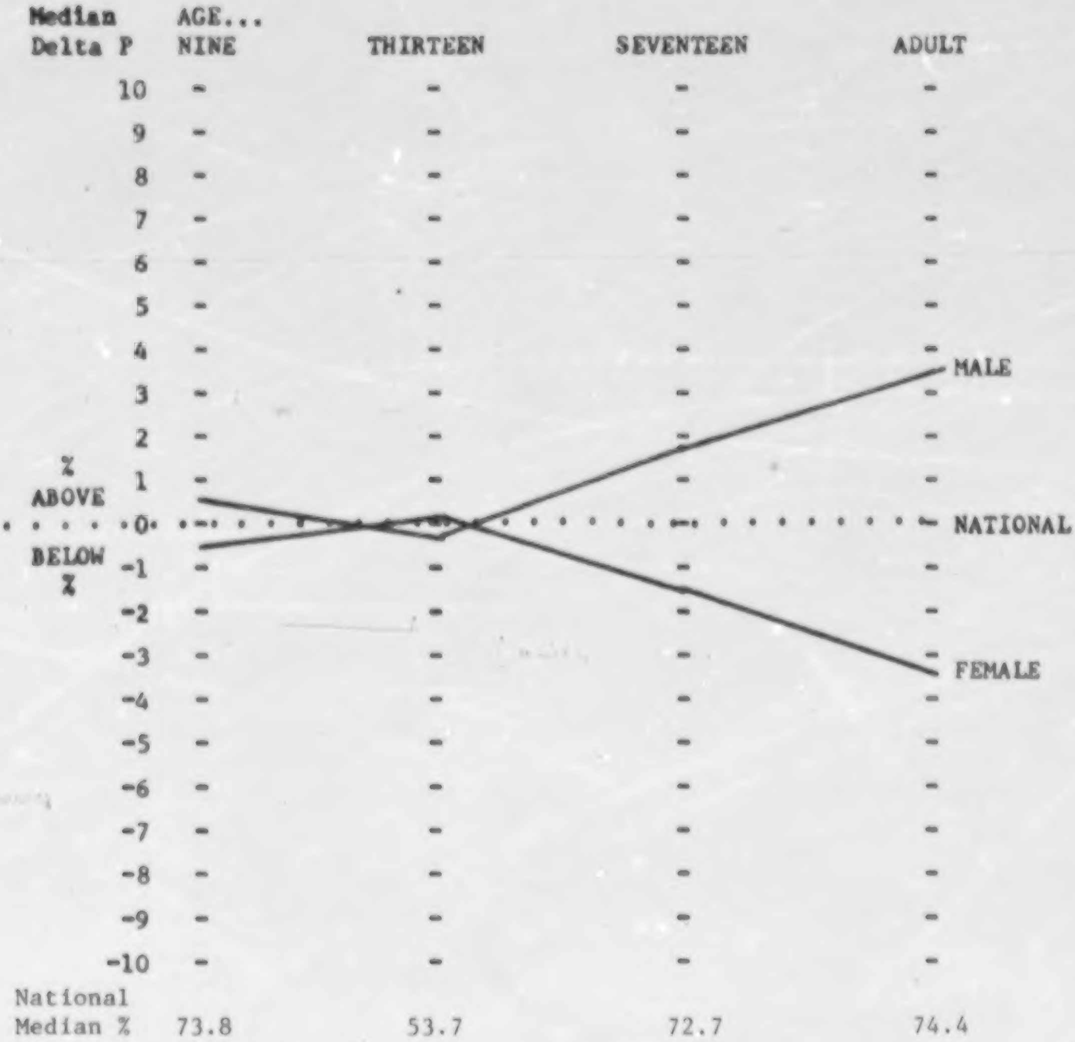


EXHIBIT 17. Knowledge of Institutional Structures - Sex - Social Studies 1971-72

C. Could you vote for both Davis and Merwin for councilman?

(National % 73.6) Yes
(Female .35; Male -.37) No
I don't know

D. If you were registered as a member of the Democratic Party, could you vote for Laski for tax assessor?

(National % 62.6) Yes
(Female -2.78; Male 2.93) No
I don't know

E. Could you vote for both Kirk and Jones for senator?

(National % 90.1) Yes
(Female .29; Male -.31) No
I don't know

Even in this set of questions where 17-year-old females scored their only significant advantage (part B), they suffer two disadvantages, one of which is significant (part D). On part D, which received the lowest national percentage of any part, the answer cannot be deciphered solely from information on the ballot. Other information about general elections must also be known. Perhaps females are a bit better in figuring out the ballot puzzles but are less informed on election rules.

At the adult level, the differentiation of males and females on the criterion of this Structure color scheme is complete. The male adults are above the national percentage on every one of the 21 exercises included; the female adults are below on all. All but 3 delta p-values out of the 42 registered for the two groups are significant. The lowest advantage scored for males is 2.6%; the greatest is 19.2%.

Males and females are most nearly alike in their responses to the voting exercises cited above and on two questions about the federal government. One of these is unreleased; the other asks which level of government could raise mail rates. They are most different on four detail questions about the federal government. Two of these are unreleased. The others ask which branch of government can declare a Congressional act unconstitutional and how a presidential candidate is nominated.

These results may suggest that the school and the society cooperate in the depoliticization of females in this country. As they grow older, males display an increasing advantage over females on knowledge about political matters and political structures.

*Rights and Duties.* Exhibit 18 depicts male and female median delta p-values for the color scheme on Rights and Duties across three age levels of 13, 17 and adult. Age 9, as we have said before, has only one exercise in this theme and, therefore, is not included here.

The median delta p-values for age 13 suggest that males and females are much alike on the 10 exercises summarized. A review of the delta p-values for individual exercises seems to confirm this judgment. The females have 7 out of 10 exercises above the national percentage of success, but none more than 2.1% above. Since all male/female distributions in these data are inversions,* the males necessarily have seven exercises below the national percentages but none more than -2.0%. The two greatest female advantages are on two unreleased exercises about freedom of speech and religion and due process. At the other end of the distribution, the males register a 6.2% advantage on an exercise that asks,

> Should a newspaper or magazine be allowed to publish something that criticizes an elected government official?
> Yes
> No
> Undecided.

At age 17 the median positions of the two groups are reversed with males being slightly above females. The dispersions are greater at age 17, however. Female advantages on five exercises in which the two groups are significantly different range between 3.3% and 9.5%. The females are significantly more successful on exercises referring to race as a factor in employment and freedom to criticize military actions and on a two-part exercise that asks,

> A. Should a congressman pay attention to the opinions and concerns of people whose views are different from those of the majority?
> Yes
> No
> Undecided
>
> B. Please explain any answer you selected.
> [Written answers judged acceptable or unacceptable].

Males, on the other hand, did significantly better on exercises referring to the right to picket to show dissent, atheists holding office and newspapers criticizing public officials.

While the difference in median delta p-values between adult males and females is only about 3.5%, the variance on individual exercises is considerably greater than for either 13 or 17-year-olds. For example, males scored a 15.3% advantage over females on the question whether persons who do not believe in God should be allowed to hold public office. Fewer than half the women sampled were willing to have an atheistic public official. The males also have about a 12% advantage over females on questions whether persons should be allowed to picket in protests against rock festivals and alleged police brutality. Over half the women (57%) would allow the picketing of rock festivals, but fewer than half (46%) would allow the picketing of police stations. While two thirds of the men would allow picketing against the festival, only a slight majority would permit picketing

---

*This inversion phenomenon is easily explained given the fact that the male and female samples are roughly equal in number in all groupings and, of course, given the necessary inclusion of everyone assessed in one or the other of the halves. Thus, the national percentage of success on an individual exercise would lie midway between the males' percentage of success and the females' percentage of success. When converted to delta p-values, a male + value would be matched by an equal female - value or vice versa. When delta p-values are ordered in a high-to-low distribution, the inversion of scores and exercises occurs. A similiar relationship is evident in black and white distributions, but the difference in numbers in the two samples and the presence of the "other" category make the inversion imperfect.

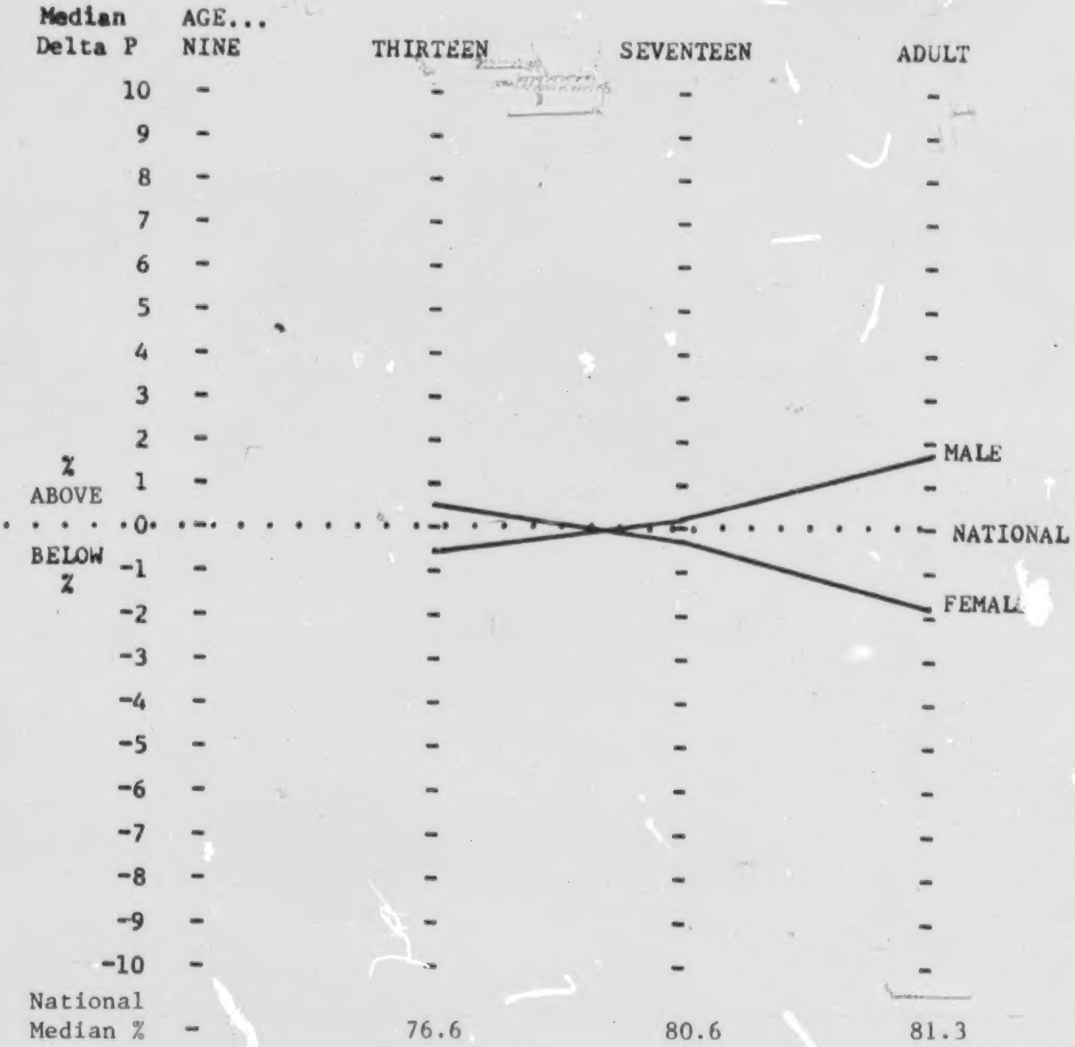


EXHIBIT 18. Knowledge of Institutional Rights and Duties - Sex - Social Studies 1971-72

against reported police brutality. Most adults (77.5%) believe that persons should be allowed to assemble in a public place to make demands known. Males, however, are 10% more likely to support such actions.

Females, on the other hand, post some advantages over males. They have about a 6.7% advantage over the men on a question about due process in a given situation and are slightly more likely than men to reject race as a factor in employment. Differences this small could, of course, be due to chance.

It would seem on the whole that males display a greater knowledge of or greater commitment to democratic rights and duties than females. This male advantage emerges between ages 13 and 17 and increases through young adulthood. One is tempted to conclude that females are socialized in this society to be less interested in, to know less about and possibly to be less committed to certain political values than males. This conclusion seems consistent with common observation. The politicized female still is an anomoly in this society. Schooling likely contributes to the depoliticization process. Schools, however, are only one factor.

## Race

Exhibits 19 and 20 display the results of these Institutional exercises subdivided into Structures and Rights and Duties color schemes and analyzed by black and white races. A third category of persons classified as "other" — i.e., presumably neither black nor white — is excluded from this analysis because of the few numbers of persons in the category and some inconsistencies in the identification of its members.

*Structures.* Exhibit 19 displays the median delta p-values for the two racial groups across the four age samples for the exercises classified as Structure.

The variance between blacks as a group and whites as a group on Structure-related questions is apparent in Exhibit 19. The median delta p-values are between 15 and 24 percentage points apart. While half of the individual exercise results are closer together than these medians, no individual exercise result for the two groups is closer than 6.6% at age 9, 2.8% at age 17 and 10.3% at the adult level. Only at age 13 does the black group score higher than the white group on individual exercises. There the blacks have advantages on two exercises of 2.9% and 8.5%. At their greatest differences on individual exercises, whites score advantages of 22% at age 9, 33% at age 13, 30% at age 17 and 39% at the adult level.

As in the previous displays of the overall results from all exercises in Exhibits 3, 4 and 5, the confirmed picture is that of two groups significantly and dramatically differentiated by their performances on these NAEP exercises.

At age 9, the two groups approach each other the closest on a question that asks,

> In a court, which one of the following has the job of making sure that the trial is fair and run according to the rules?
>
> The judge
> The lawyer
> The jury
> The person on trial
> I don't know.

Nearly three fourths of all 9-year-olds could answer that question correctly. Whites are only 1.6% above the national percentage while blacks are only -5.0% below. Black and white 9-year-olds come nearer equally sharing this information than any other in this color scheme. At the other extreme, only one third of all 9-year-olds know that health departments inspect restaurants. When divided by race, about 40% of the white children answered the question correctly, while only 17.6% of the black children know this fact.

Explanations for such results tumble out rapidly. As a group, 9-year-olds are not especially aware of the ordinarily unsung governmental function of restaurant inspection. Just as likely, they are not sensitized to culinary cleanliness or possibly to restaurants themselves. Beyond this, black youngsters probably frequent restaurants less than white youngsters.

On the other hand, blacks frequent courtrooms in this country as trial principals far beyond what their population proportion would predict. Furthermore, television dramas offer recurrent though usually distorted views of courtroom scenes in which judges mainly play referees between wily lawyers.

In another instance, over 60% of the white 9-year-olds could identify the mayor as the head of a town government, while only 45% of the black 9-year-olds could do so. Cultural explanations of this 15% variance are harder to come by, but one possibility is that blacks in this country tend to be either city or farm folk. Towns are predominantly white. Thus, probably a smaller proportion of black than white 9-year-olds in the sample would be town residents. However, we do not have that kind of information on the composition of these samples.

At age 13, the Structure color scheme is assessed by means of 16 exercises. The range of successes over these exercises for white 13-year-olds goes from -1.0% below the national percentage to 4.5% above; the range for black 13-year-olds goes from -29.0% below the national percentage of 7.5% above.

On one unreleased exercise for age 13 concerning a function of a major department of the federal executive, black 13-year-olds achieved an advantage of 8.5% over the whites. It is also the exercise on which black 17-year-olds did best while white 17-year-olds did very poorly. In both cases, since the whites, who outnumber the blacks in the samples by 6 and 6.5 to 1, did badly, the national percentages of success are quite low. Only 27% of all 13-year-olds and 38% of all 17-year-olds answered the exercise correctly. The result may have very little meaning, actually, since the national percentage is at the chance level of performance at the 13-year-old level.

On three related questions about government services, black and white 13-year-olds are extremely divergent in their successes. Students were asked "Which level of government (federal, state or local) would be *most* likely to pass" the following:

> . . . an act to raise the rates for sending letters through the mail?
> . . . an act to lower taxes on goods coming into the country?
> . . . an act to increase garbage collection services?

The national percentages on the three questions are 71.7%, 73.4% and 77.0%. The white 13-year-olds are above the national percentage in each case by 3.6% to 4.5%, but the blacks register deficits from -21.0% to -23.9%. A bare majority of black 13-year-olds could answer the questions correctly, while over three quarters of the whites responded correctly.

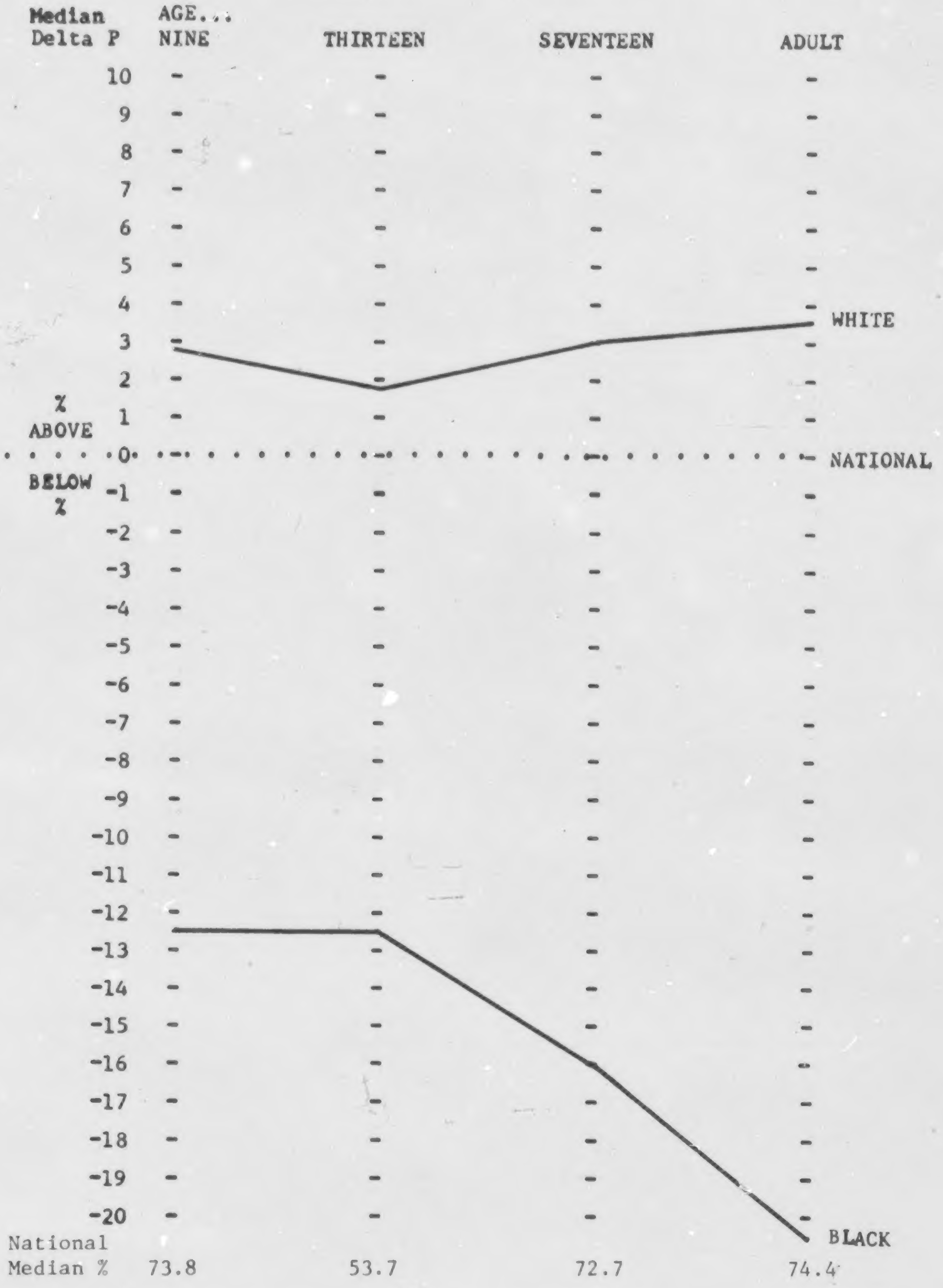


EXHIBIT 19. Knowledge of Institutional Structures - Race - Social Studies 1971-72

A whole set of cultural effects is probably operating here. It would seem gratuitous to point them out. Whether they are sufficient to explain the 25% disadvantage of blacks is, of course, the central problem.

We will mention one other exercise, which is the locus of poor performances of all the samples in the assessment. Age 13 respondents scored their lowest national percentage of success on this exercise. They were asked,

> The presidential candidate for each major political party is formally nominated by which one of the following?
>
> The Senate
> A national primary
> A national convention
> The House of Representatives.

Only 17.0% of the 13-year-olds could answer the question correctly. Whites managed to make that 17.9%; blacks scored a bare 10.8%. Presidential campaigns seem not to be a major interest of American 13-year-olds students, black or white.

At age 17 the median advantage of whites over blacks increases to 19%. This difference reflects a range of differences on individual questions from as little as 2.9% to as great as 32.1%. All of these differences in delta p-values are white 17-year-old advantages over black 17-year-olds. Thus, on all but one exercise whites scored significantly above the national percentages; on all exercises black 17-year-olds scored below the national percentages, and all but one are significantly below. The black and white 17-year-olds responded alike — i.e., neither is significantly different than the national percentage for all 17-year-olds — on an unreleased exercise about the functions of a department in the federal executive. Albeit, that result is by far their lowest national percentage of success; only 37.7% of all 17-year-olds marked it correctly.

Black and white 17-year-olds are most different in their responses to two questions about the federal government and one question about a function of local governments. On the two of these that are unreleased, one of the federal government questions and the local government question, 53% and 88% of the white 17-year-olds marked them correctly while only 28% and 57% of the black 17-year-olds did so. The released federal question asks how a presidential candidate is nominated. While barely a majority of whites could answer this question, only one in five of the blacks marked it correctly, less than could be expected on the basis of guessing.

Among adults the differences between blacks and whites widen still further. While there is general gain in adult national percentages of success as well as in all adult raw p-values for both blacks and whites over comparable values for 17-year-olds on these Structure questions, all of which are shared by adults and 17-year-olds, the blacks tend to gain fewer percentage points and less consistently than whites. All white adult delta p-values are significantly above the national percentages and all black adult delta p-values are significantly below the national percentages. Within their ranges of typicality, black and white scores on individual questions vary from 19 to 30 percentage points, always with a white advantage. Where they approach each other in performance — i.e., on an item where blacks do their atypically best and whites their atypically worst — the whites post only a 10.4% advantage. This unreleased exercise has to do with the armed forces. At the other end of the scale, on the question about national political conventions, whites have a 39.2% advantage. Only 26.2% of the black adults could answer the question, a guessing score. Nearly two thirds of the white adults know that presidential candidates are nominated by national conventions.

Three of the four questions that are the loci of atypical performances by both blacks and whites are also among the five easiest questions for adults. The national percentage for each of the five exercises is above 90%. On one of these, the ballot question where respondents are asked if they could vote for both Democrat Kirk and Republican Jones for the single Senate seat, 80% of the blacks and 92% of the whites could answer correctly. On another, 97% of the whites and 82% of the blacks noted that the federal government could raise mail rates.

On two other questions, blacks do comparatively worse. Only 71% of the black adults as compared to 95% of white adults could relate the federal government with the control of tariff rates. At the local level, only two thirds of the black adults connect garbage collection with local government; 95% of the white adults know this common local government function.

While there is improvement in the performances of both blacks and whites from the lower to the higher age samples on the shared questions, it seems undeniable that the relative disadvantage of the black samples increases with increasing age. Schooling may contribute to the ability of all persons to respond to questions such as these, but it is not effective in reducing the performance gap between these two racial groups in our country. That such a gap can be described racially is evidenced in this NAEP assessment and in most other evaluative efforts that enjoy mainstream sanctions. That such gaps become defined racially is the unfortunate and unwarranted consequence.

*Rights and Duties.* Exhibit 20 displays the group effects of the three black and white age levels for which there are exercises classified as measuring a mix of knowledge of and commitment to some Rights and Duties associated with some of our social institutions.

The relative advantage of white respondents over black respondents ranges from 8.5% at age 17 to 17.9% at the adult level. The apparent gain in black performances noted at age 17 on this set of exercises is also reflected in the Southeast group on the regional displays and in the low metro group in the STOC exhibits. The Southeast and the inner city are likely locales of the majority of black respondents in these samples.

Within their ranges of typicality, black and white 13-year-olds differ from each other from 7.2% to 17.4% on individual exercises. Within this group of questions to which both groups responded typically is the question,

> Should race be a factor in hiring someone for a job?
>
> Yes
> No
> I don't know

"No" is the response of 90.5% of the white 13-year-olds and the response of 83.3% of the black 13-year-olds.

The two groups are actually closer together and unexpectedly more settled on a question that asks,

> Do you think there should be laws against acts of vandalism such as destroying a statue?
>
> Yes
> No
> Undecided.

Nearly 95% of the blacks and 93% of the whites approve of

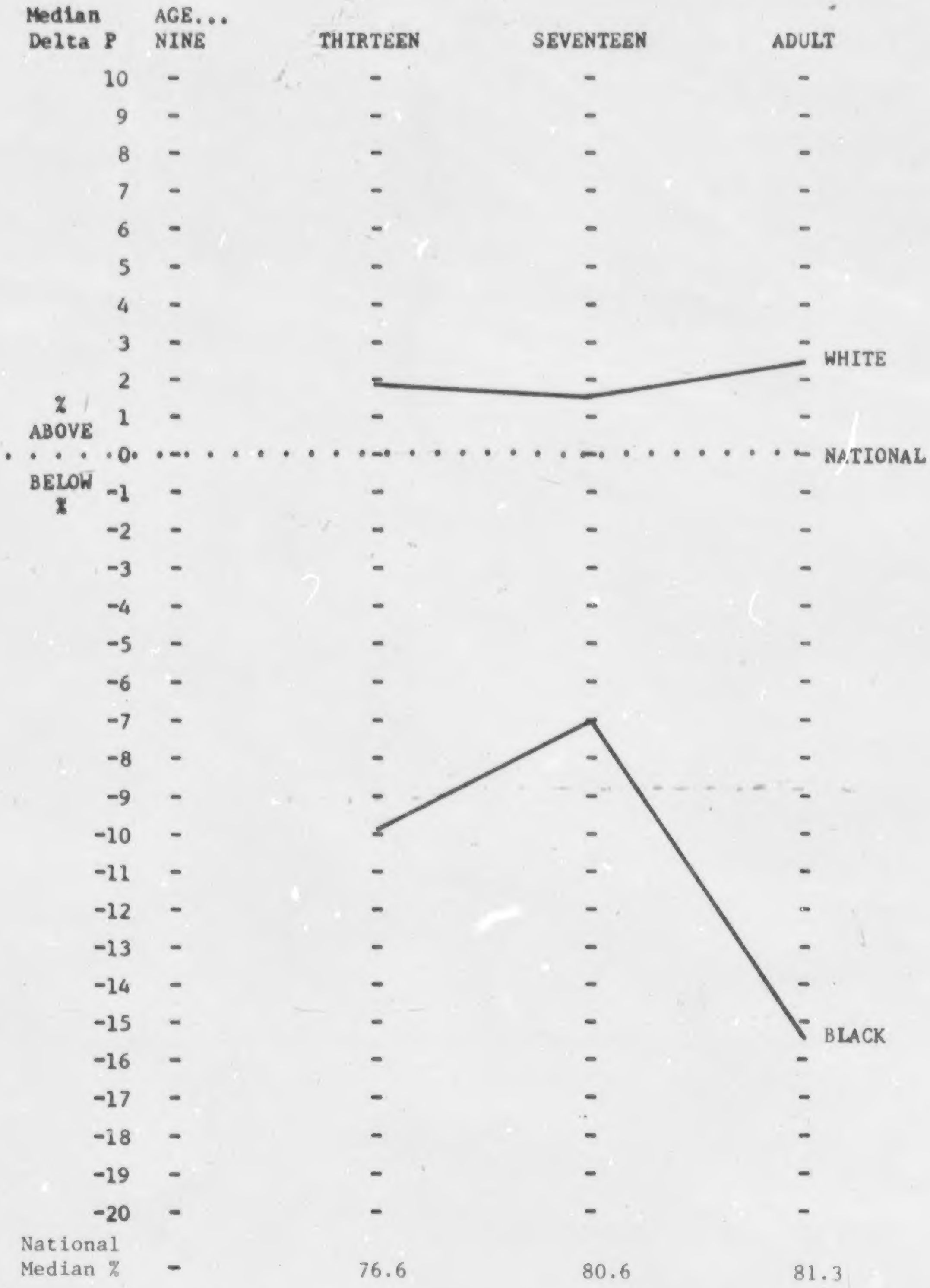


EXHIBIT 20. Knowledge of Institutional Rights and Duties - Race - Social Studies 1971-72

such laws. This is the only question on which black 13-year-olds post an advantage over whites. A contrast in performance and in content is found in an unreleased exercise referring to religious freedom. While 81% of the whites supported this value, only a bare majority of blacks did. To suggest that the responses to the two questions, especially among blacks, demonstrate a disparity between human and property values is probably too facile.

At age 17, the Rights and Duties color scheme is defined by 21 exercises. Nine questions are shared with the 13-year-old group and all but one are shared with the adult group. As in all other distributions featuring black and white respondents, the dispersion of the whites' delta p-values is quite compact, running from -1.5% to 4.8% for a total range of only 6.3 percentage points, while the dispersion of the blacks' delta p-values is broad, running from -24.5% to 7.6% for a total range of 32.1 percentage points.

The two groups are practically alike in their responses to the question involving racial discrimination in employment and an unreleased question on age discrimination. On the racial discrimination exercise, the two groups' performances are within 0.8%; on the age discrimination question only 1.4 percentage points separate the two groups' results. The national percentages for the questions are 93.8% and 89.3%.

Black and white 17-year-olds are most different on a question involving the freedoms of speech and religion and a question that asks,

> Should a person who does not believe in God be allowed to hold a public office?
>
> Yes
> No
> Undecided

The groups' responses vary 21.2% on the first and 29.1% on the second. About 60% of the black 17-year-olds appear to be supportive of religious and press ~~freedoms in the~~ first, but only about 38% are supportive of religious freedom in the second, as compared to 82% and 67% of the white 17-year-olds.

Also among the most varied responses are those related to the question,

> Do you think people should be allowed to picket the holding of a rock festival as a protest against it?
>
> Yes
> No
> Undecided
> [illegible] response.

A small majority of 52.7% of all 17-year-olds would support this sort of action. As a group, whites are slightly more willing; 56.5% would approve. However, among black 17-year-olds, only 32.8% are approving. The contrast between blacks and whites is complicated by the fact that just half the white 17-year-olds approve of the picketing of a police station to protest reported police brutality; but 59.2% of the blacks would approve. Roughly speaking, almost twice as many black 17-year-olds would allow a picket protest against alleged police brutality as would allow a picket protest against a rock festival. The effect of social experience on the way persons respond to presumably similar situations, picketing, in this case, seems well demonstrated. Perhaps, the effect of social experience on the way persons respond to NAEP test questions is better demonstrated.

Adult black and whiite respondents demonstrate the same patterns on Rights and Duties exercises as do the 13-year-old and 17-year-old samples. All but two delta p-values of whites are above the national percentages and all but two delta p-values of blacks are below the national percentages. The median delta p-values in Exhibite 20 suggest further that on Rights and Duties exercises, as was also noted earlier for the Structure exercises, black and white adults differ more from each other than do either the 13 or 17-year-old blacks and whites.

On three exercises in this color scheme, black and white adults are most alike, i.e., neither group is significantly different than the national percentages of success. It is on these exercises that the Whites produce their lowest delta p-values and the blacks produce their highest. The two released questions refer to the factor of race in employment and picketing a police station. Over 90% of both groups believe race should not be a factor in employment and just over 50% of each group believe that persons should be allowed to picket against brutality. The radicals among the adults are those persons classified as "other" who are not included in this report. They posted an impressive advantage of 16.9% on the police-picketing exercise.

The adult groups are most different on three released exercises where the whites are from 3.7% to 5.0% above the national percentages and the blacks are from 20.4% to 29.9% below. The three exercises, previously referred to in other sections of this chapter, concern atheists holding public office, picketing a rock festival and the Supreme Court decision on religion in public schools. The national percentages on these questions are 55.6%, 63% and 52.3%. The principles involved in these questions are either not too well known or are not too well regarded by this sample of adults or are opposed by other values unknown.

According to these results, the principles are also not equally known or regarded by these black and white adults. In the following exercise, 57.4% of the white adults sampled recognize the principle of separation of church and state, but only 22.4% of the sampled black adults know the grounds for the famous decision.

> The Supreme Court ruled that it is unconstitutional to require prayer and formal religious instruction in public schools.
> Which one of the following was the basis for its decision?
>
> The requirements violated the right to freedom of speech.
>
> There was strong pressure put on the Supreme Court by certain religious minorities.
>
> Religious exercises violated the principles of the separation of church and state.
>
> Every moment of the valuable school time was needed to prepare students to earn a living.
>
> I don't know.

"I don't know" presumably functions differently than a substantive distractor in such a question; however, given the right assumptions about the four distractors in the exercise, guessing by the entire black adult population in this sample could be expected to yield the same result they produced. Neither do the whites earn accolades for their performance. Chapin's team, in another inquiry into this assessment, estimated that at least 60% and possibly more than 80% of the adults should be able to answer this question.

Besides the chance results of guessing, conjectures as to why these adults performed so poorly here are myriad. They would have to include, it would seem, the possibility that these adults, in the absence of recall knowledge about the case, respond more readily to suggestions of political practicality than to moral principle and, therefore, project the same posture to others, including Supreme Court justices. It is far more practical to suppose that the court yielded to pressure — not a bad guess, really, given the political nature of that institution.

In another of these differentiating exercises, about 60% of the white adults appear willing for a person who does not believe in God to hold public office. That majority is not likely to get an atheist elected in many communities, however, for only about 35% of the black adults are so inclined. The anomoly is that blacks, who have been extraordinarily influential in expanding civil rights relative to race, appear not to extend these principles to other realms such as religious belief.

The last of these exercises also suggests some contrasts. Some 67% of the white adults would allow pickets to protest a rock festival. Only 39% of the black adults would do so. The contrast between the groups anent rock festivals, which may reflect different musical tastes, is less telling than the contrast with these groups' previously cited willingness to picket against reported police brutality. Just over half of both groups would allow anti-police pickets. Thus, to draw the lines of perspective, the blacks are more willing to allow active protesting of police brutality than they are to allow active protesting of rock festivals. Whites, on the other hand, seem more concerned about the festivals than about the brutality. Perhaps the difference has something to do with whose ox is gored.

### Parental Education

In this section, the results for the four age levels in the two color schemes are analyzed across four levels of parental education. The median delta p-values used to characterize the different groups are shown in Exhibits 21 and 22.

*Structures.* Exhibit 21 displays the results in the Structures theme for ages 9, 13, 17 and adult classified according to whether neither of the parents of the respondents went to school beyond the eighth grade, at least one parent started but did not finish high school, at least one parent graduated from high school but did not go beyond that, or at least one went on for some kind of training after graduating from high school. The two exhibits attempt to show the relative relationships between the respondents' answers to these questions and the varied educational exposures of their parents.

The displays strongly suggest that the four age samples can be divided into discrete populations according to the different levels of parental education. The relative positions of the median delta p-values are hierarchical and consistent. The higher the educational attainment of the parents, the better the respondents perform on the set of questions classified by this theme for each age level.

At age 9, the sample is almost dichotomous with the no high school and some high school groups producing all results on individual questions below the national percentages: 8 of the two groups' 10 delta p-values are significantly below. In contrast, all of the 10 delta p-values of the graduated high school and post high school groups are above the national percentages; 8 are significantly above.

The four age levels come nearest performing as similar populations on the question that asks who is responsible for a fair trial. Only the post high school group answered this question at a level that is significantly above the national percentage of 73.8%. The post high school success of 77.6% has an advantage of 10.7% over the lowest no high school group.

The post high school group's advantage over the no high school group is 18.0% on the "town mayor" question and 16.5% on the "restaurant inspection" question. On both of these questions the post high school group is significantly above the national percentages while the no high school group is significantly below. Only 44% of the no high school group know that the mayor is the head of a town government and only 27.2% know that health departments inspect restaurants. Few 9-year-olds in any category know this health information. Only 44% of the post high school group marked the question correctly.

The dichotomy between the post high school and graduated high school groups and the some and no high school groups observed at the 9-year-old level seems not to hold at age 13. For one thing, though not detectable in the median values shown in Exhibit 21, three of the groups' scores are far more disperse at age 13 than at age 9. At age 9, the spreads of the four age categories on five observations are 5.9%, 8.6%, 4.1% and 4.4%. At age 13, the spreads on 16 observations are 26.5%, 15.1%, 4.2% and 13.4%. The anomoly is the graduated high school group with a dispersion of only about 4.0% in both age samples. There seems no self-evident explanation for the anomoly. The graduated high school group simply performed close to the national percentages on all the exercises. The some and no high school groups hover near the national percentages on perhaps one third of the exercises but plunge 16% to 24% below on others — though not necessarily on the same ones. Meanwhile, the post high school group soars above the national percentages from 2.4% to 13.1% on all but one exercise.

The groups performed alike only when everyone appears to have been guessing. For example, an unreleased exercise asks for the identification of a federal executive department by its major function. All scores converge to the national "guessing" percentage of 26.7%.

The greatest differences in the groups are registered for the two exercises that ask for the identification of the level of government — i.e., federal, state or local — most likely involved in lowering taxes on incoming goods and in increasing garbage collection. The post high school group earned its greatest advantages on these questions with successes of 13.1% and 11.0% above the national percentages while the no high school group, in contrast, suffered its greatest disadvantages with successes of -16.6% and -23.7%. The graduated high school group, which deviates from the national percentage but little on any question, has a delta p-value of -3.15% on one and -2.35% on the other. The some high school group has one of its worst scores on one, -11.9%, and one of its best on the other, -1.3%.

Generalizations at the 13-year-old level, other than those suggested above, are elusive. While overall the 13-year-olds do not perform well on this set of questions, in general they do somewhat better on exercises whose contents have recurrent reinforcement in day-to-day experience. For example, practically none know how a presidential candidate is nominated (17% for all; from 12% to 20% by group); but most know how men get into the armed forces in wartime (82% for all, from 70% to 90% by group). There are results, however, that make this generalization less than firm. For example, more no high school respondents know that a

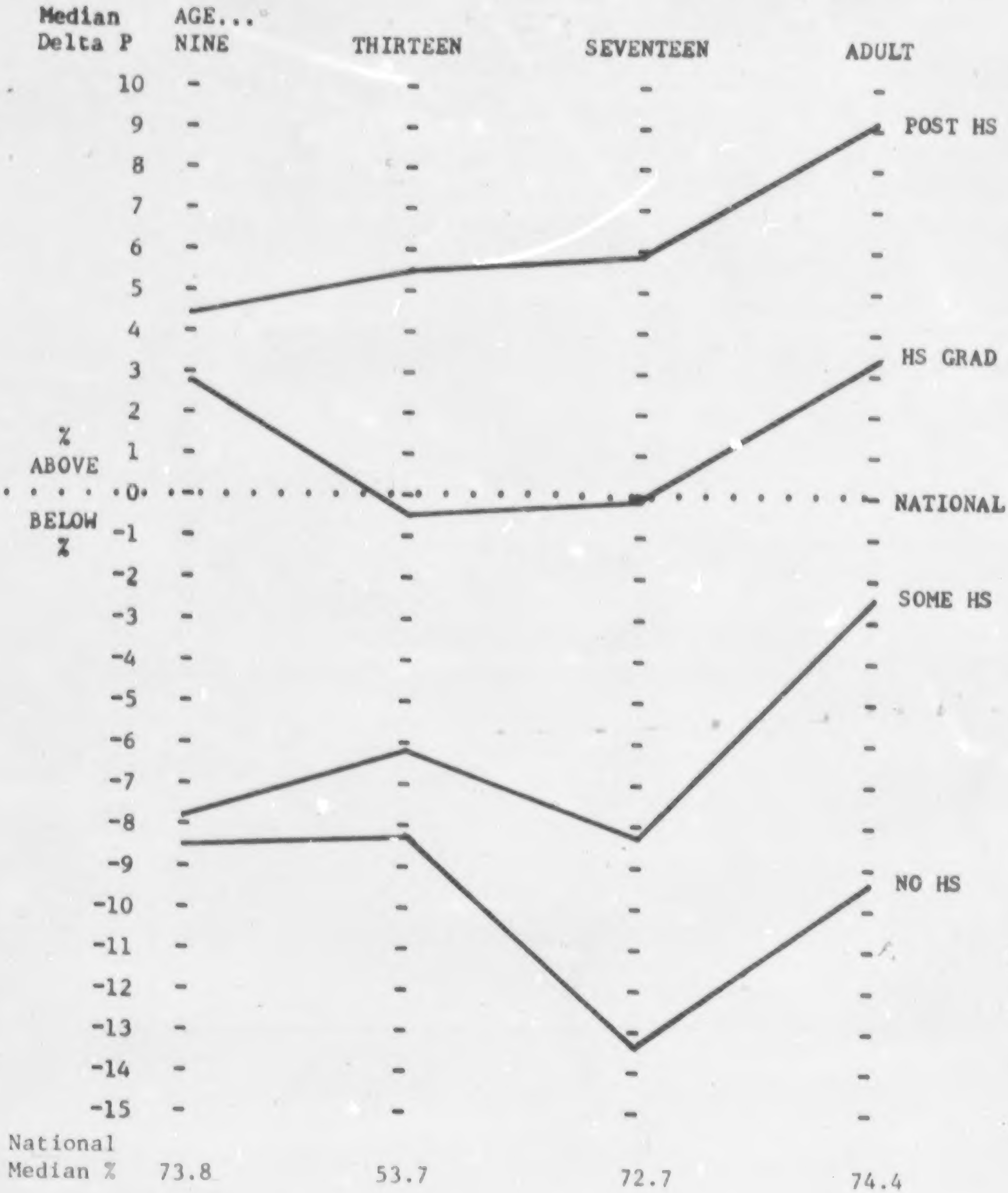


EXHIBIT 21. Knowledge of Institutional Structures - Parental Education - Social Studies 1971-72

senator usually is elected rather than appointed (70%) than know that local governments collect garbage (52%).

At age 17, the groups appear to perform much as they do at age 13. All post high school 17-year-olds' delta p-values are above the national percentages and all but one are significant; all no high school delta p-values are below the national percentages and all but one are significantly below. The some high school 17-year-olds look a bit like a replica of the no high school group, but with delta p-values elevated about 5 percentage points. Nonetheless, all but one of these values are below the national percentages; all those below are significant. The 5% difference is accurately reflected in the median delta p-values in Exhibit 21. The post high school 17-year-olds are something of a reversed image of the some and no high school groups with their best relative performances on exercises on which the other groups performed most poorly. The dispersions of these three distributions are also quite large. The no, some and post high school groups' spreads, in that order, are 20.7%, 22.6% and 11.3%. The graduated high school group maintains its anomolous character. Its delta p-values are all within about ±3% of the national percentages, half above and half below. The list of 21 exercises, ordered from the group's best to poorest performances, does not resemble the lists of the other groups when ordered by the same rubric. An acceptable explanation is no more apparent for the 17-year-olds than for the 13-year-olds.

One explanation, which lies somewhere between a pure conjecture and a random guess, is that the group whose parents graduated from high school but never went beyond high school has the strongest orientation to school. It becomes the norming group on these exercises, which also reflect a strong taint of schooling. The other groups, those whose parents were less or more successful in school, deviate in different ways from the norm.

There seems to be some evidence for this hypothesis. The best example of it relates to the exercise on the nomination of presidential candidates by national political conventions. The national percentage for this question for all 17-year-olds is 48.5%. The graduated high school group attained 48.6% on this question. Both the some and no high school groups turned in their worst performances on this question with delta p-values of -18.8% and -22.6%, respectively, while the post high school group turned in its best performance with a delta p-value of 12.1%.

In another instance on an unreleased exercise concerning school governance, the graduated high school group is about 1.0% below the national percentage and within 0.9% of its own median delta p-value for all of these exercises, while the no and some high school groups produced their best delta p-values of -1.9% and 3.8%, respectively, and the post high school group produced its worst delta p-value of 0.7%.

At the adult level, differences in the groups continue to manifest themselves. Inequalities in performance in this assessment associated with different family environments with respect to parental education are not erased by time. Having been reared by parents who did not attend high school appears to disadvantage adults with respect to this assessment, in much the same manner that it appears to disadvantage others of similar backgrounds of school age. Furthermore, while there is some reason to infer from these data that there is a slight improvement in test performances of this kind by persons after 17 years of age, there is more basis for inferring that the important gains are made during the school ages of 13 and 17. These inferences can be made largely from the median national raw p-values shown in Exhibits 3, 4 and 5 in the early part of this chapter. The evidence here relative to Structure information suggests that among groups all such improvements are parallel. No group — here, for example, groups categorized by parental education — escapes its past.

An example of the kind of improvement referred to is in the question about a presidential candidate's nomination. Less than half of all 17-year-olds could answer this question, but nearly 60% of all adults know this function of national conventions. That would count as improvement if it is more than a chance gain. The persistent inequalities are shown by the fact that as adults, still less than half (44%) of the persons whose parents never went beyond eighth grade know how presidents are nominated. About 57% of the some high school group know this fact along with about 64% of the graduated high school. In contrast, 80% of the group whose parents went on to school after high school marked the question right.

Some things, it seems, nearly all adults know. That the federal government carries the mail is known by 95% of all adults. Even here, however, group differences persist. The no high school group posts a high 92.3% on this question, but the post high school group hits a cracking 99.2%.

A phenomenon of these adult data when categorized by parental education is the uncannily accurate ordering of the exercises by difficulty in the post high school distribution. That is, on the hardest exercises, those which received the lowest national percentages of success, the post high school group achieved its highest delta p-scores. But on the easiest questions, where the national percentages of success were highest, the post high school group achieved its lowest delta p-values.

One's first impression is that this group did best on the hard questions and poorest on the easy questions; but that is not the case, for on the easiest question of all for the adults as a whole, the post high school group has a percentage of success of 99.3%. That means only three persons in the whole population missed the question. The fact is that the post high school group did well on all the questions. When the national percentage was high, there was simply less range on top for the post high school group's advantage.

A second hypothesis is hazarded. The most consistent, effective and long-lasting influence on the ability to respond to questions of this kind is the educational environment of the home. Another way to put it is that parents who are successful in school have children who are successful in school. That assumes, of course, that the ability to answer these questions is related to or is more generally expressed as the ability to succeed in school. Persons who are successful in schoolwork are successful partially, at least, because they develop interests in the things schools stand for and emphasize. To some extent, perhaps to a great extent, such school-related interests are retained into adult life and are perpetuated as interests in the lives of children, "visiting," as it were, "the iniquity of the fathers upon the children to the third and the fourth generation . . . ."

*Rights and Duties.* Exhibit 22 displays the median delta p-values for three age levels for the exercises in the Rights and Duties theme as categorized by parental education. Age 9 is not represented because of too few exercises in this theme.

In general terms, the relationship of parental education to percentages of success in these thematic arrays is consistent with what was found in the Structure thematic distribu-

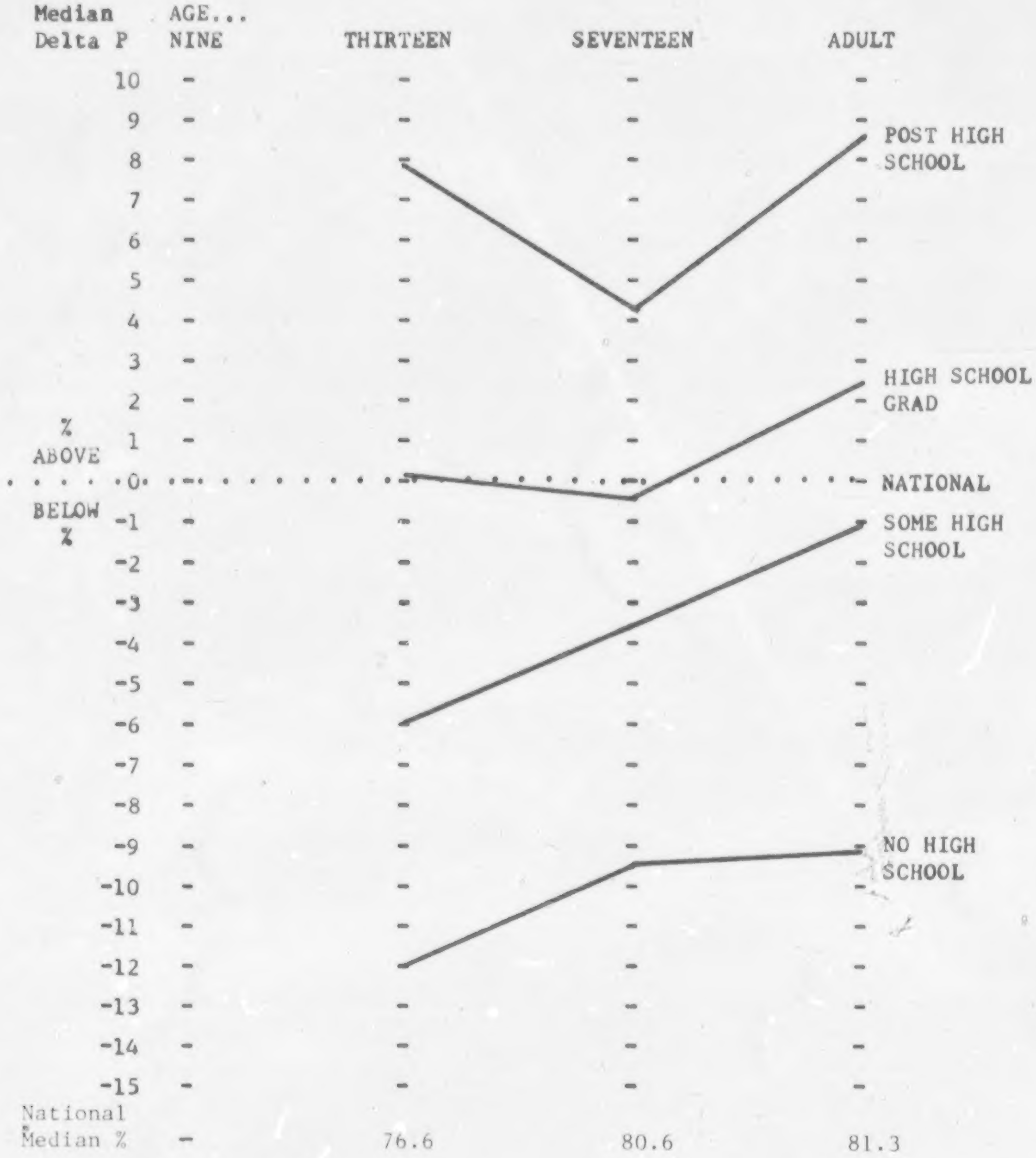


EXHIBIT 22. -Knowledge of Institutional Rights and Duties - Parental Education - Social Studies 1971-72

tions. Sorting on the four levels of parental education appears to result in four distinct populations that perform consistently in this theme at all age levels. The post high school group scores advantages consistently; the other groups consistently score relatively lower successes.

As we have emphasized before in reviewing this theme, responses in this theme may carry a heavier value burden than those in the Structure theme. Knowledge of Structure questions are based largely on factual recall while the questions classified as Knowledge of Rights and Duties are an uncertain mix of fact and value. To some degree, the delta p-values reported in this theme reflect a rating and ranking of preferences associated with our democratic institutions. In other respects, they reflect a knowledge of the principles and rules by which our institutions are governed. To say, for example, that race should not be a factor in employment may indicate one's preference in the matter or it may indicate a knowledge of the law of the land, which explicitly prohibits racial discrimination in certain jobs and implicitly discourages it in most others.

Among 13-year-olds, a distribution pattern similar to that noted earlier for Structure is evident. Dispersions of the no, some and post high school groups are fairly wide — 21, 14 and 11 percentage points, respectively — while the dispersion of the graduated high school group is narrow, only 5.5 percentage points. Furthermore, all the no high school delta p-values are below the national percentages, as are all but one of the some high school group's scores. In contrast, all post high school values are above the national percentages. The graduated high school group's delta p-values are half above and half below.

On only one question did all groups respond in the same way; 93% of all groups, ±0.4%, agree that there should be laws against vandalism. On all other exercises there is significant variance by one or more of the groups. No other question received so high a national percentage of success. A sense of property rights still appears to be alive and well, at least among these 13-year-olds.

At the opposite end of the success scale, only 48.9% of all 13-year-olds responded affirmatively to the following question:

> Should a newspaper or magazine be allowed to
> publish something that criticizes an elected
> government official?
>
> Yes
> No
> Undecided.

This question also produced considerable variance among the groups. The post high school group is 10.5% above the national level on the question while the graduated high school group, its nearest competitor, is -3.9% below. The some and no high school groups are far below the national percentage at -14.5% and -13.1%. While 9 out of 10 of the respondents whose parents either did not go or did not graduate from high school believe that vandals should be punished, only one out of three believe government officials should be publicly criticized in the media. In addition, fewer than half of these groups would allow an atheist to hold a public office. On this question, over 60% of the graduated high school group and 67% of the post high school group would permit a person with such a belief to hold office.

One is tempted with such results to apply labels of liberal and conservative to the groups. However, conservative is hardly the proper appellation for a group two thirds of which would not allow newspapers to criticize the government; nor is liberal an especially appropriate way to refer to a group one third which would apply a religious criterion to public service. The application of a value-conflict model to these data would likely surface a number of both inter- and intragroup conflicts that warrant some attention.

No great liberal group shows up at age 17 either; though a tendency in that direction can be inferred for the post high school group on some questions, if certain assumptions are made about the meaning of liberal, the meaning of the questions and the meaning of the responses to the questions. For example, the post high school 17year-olds have significantly positive responses of 4% to 10% above the national percentages on questions involving public assembly, picketing police stations and rock festivals, criticizing government officials and policies and atheists holding public office. Their advantage over the some and no high school groups on these questions runs from 5% to 30%. If the assumptions can be made that a person supporting the activities listed above can properly be thought of as more liberal than a person not supporting them and that support or approval is what is being indicated by the positive marking of those questions, then it follows that the post high school group shows more of a tendency toward liberality than any other group. Admittedly, the conclusion is tenuous. It would be just as valid to suppose that the exercises measure an understanding of our way of life inasmuch as all the activities referred to in the questions are in accord with the law of the land. That the 17-year-olds achieve on these exercises national percentages of success of only 50% to 90% may more aptly suggest an imperfect acquaintance with some elements of our democratic system.

There is evidence in two other sets of exercises in this theme that suggests that the shared orientation of this 17-year-old sample is "law and order." The inference is derived from a set of questions on which there is fairly high agreement among these respondents and little variance among the groups in comparison with a set of questions on which there is relative low agreement and large variation across groups.

When asked if there should be laws against vandalism and if a crime against property would be reported by them, 92.6% and 74.4% of these 17-year-olds answered affirmatively. There is nothing in those responses that is necessarily improper, of course. Percentages that high simply suggest high concern and strong agreement. No group answered either question in a significantly different way.

In something of a contrast, when asked if they would allow picketing against a rock festival and picketing against reported police brutality, only 52.6% and 51.6% of these 17-year-olds answered affirmatively. Furthermore when analyzed by the four education categories, there is considerable disparity across the groups. The no high school group is significantly below the national percentages by -19.1% and -13.8%; the some high school group is significantly below the national percentages by -14.9% and -1.8%; the graduated high school group is below the national percentages by -3.9% and -2.8%; and the post high school group is significantly above the national percentages by 10.8% and 4.6%.

The point is that the 17-year-olds are very certain and very homogeneous on the former questions characterized here as having a law and order orientation, but they are pusilanimous and disparate on the latter questions involving matters of a different order.

At the adult level, this law and order characterization can also be supported reasonably well by comparing the adult groups' responses on the three questions that are shared by the two age samples. When asked the question involving the reporting of a crime against property, 92.4% of the adults answered affirmatively, a much higher percentage than for 17-year-olds. The graduated high school adult group is significantly above the national percentage on the question by 2.1%. No other group answered the question significantly different from the national percentage. The adults, like the 17-year-olds, are quite certain and fairly homogeneous on the question.

When asked about picketing rock festivals and police stations, 63.0% and 52.2% of the adults answered affirmatively. They are 10% more willing than the 17-year-olds in the one case and about the same as the teen-agers in the other. When analyzed by education categories, the adult groups, like the 17-year-olds, are markedly different. The no high school group is significantly below the national percentages by -11.1% and -8.1%; the some high school group is below the national percentages by -4.3% and -4.2%; the graduated high school group is above the national percentages by 5.9% and 2.3%; and the post high school group is significantly above the national percentages by 15.2% and 12.3%.

Beyond this possible orientation, there are other useful comparisons to make among the adult groups. They are also agreed that race is not a proper factor in employment. Adults are verbally accepting of racial equality in employment at the 92.3% level. There are no differences according to parental education.

At the other extreme, only 55.6% of all adults would permit an atheist to hold public office; but they vary greatly according to the educational environment of their youth homes. Only 41% of the adults whose parents never went to high school would allow a nonbeliever to hold office while 76% of those whose parents went beyond high school would do so.

Also, a bare majority of adults agree that a constitutional principle guided the Supreme Court's ruling on religious instruction and prayer in public schools. The post high school group registers a respectable 72.4% on this question, but only 39% of the no high school group agree with this response.

## Size and Type of Community (STOC)

Exhibits 23 and 24 display the median delta p-values for the two color schemes, Knowledge of Institutional Structures and Knowledge of Institutional Rights and Duties, categorized by size and type of community for the four age samples, 9, 13, 17 and adult. While all seven STOC categories are exhibited, most of the discussions will refer only to the high metro, low metro, urban fringe and extreme rural groups. The justification for this delimitation is in the contrast afforded by these groups and in their characterizations. The high metro group is mainly affluent, the low metro is largely inner city, the urban fringe is suburban and the extreme rural is primarily farm. More complete definitions are included in Appendix B.

*Structures.* Exhibit 23 shows the median delta p-values for these groups and age samples for the Institutional Structures color scheme.

Among the various STOC distributions in the 9-year-old sample, the most glaring conclusion is the gross difference between the delta p-values of the high and low metro groups. That difference, of course, is patently evident in the display of medians in Exhibit 23. Actually the high metro group differs less from the rest of the sample than does the low metro group. The high metro group is significantly above the national percentages on only two of the five exercises in the theme while the low metro group is significantly below the national percentages on four of the five exercises.

The lowest low metro dispersions of about -16% are on an unreleased question about taxes and the question about restaurant inspection. On the tax question the national percentage is 83.2%. The high metro group is a significant 5% above this level. None of the other five groups vary significantly on this question. The poorest performance by all 9-year-olds is on the restaurant inspection question. Only 36% know that this task is a function of health departments, but fewer than one out of five of the low metro group know this fact. That is probably below the guessing level.

About three fourths of the 9-year-olds know that judges are responsible for fair trials. This knowledge is shared across all these STOC groups about equally. Only the urban fringe group registers a delta p-value significantly above the national percentage, and that advantage is only 3.6%. The high metro group is actually 1% higher, but the statistical controls do not label that difference as significant for the high metros.

At age 13 there is an apparent slight narrowing of the differences between the high metro group and the low metro group as shown in Exhibit 23; but a sharper look at the distributions from which these medians are drawn do not bear this out. If anything, the two groups are significantly different in more areas; but, of course, there are more exercises at this age level. The other groups, possibly excepting the urban fringe and main big city groups, are so nearly alike to be indistinguishable; that is, none vary essentially from the national percentages.

The high and low metro groups share three exercises on which neither is significantly different than the national percentages. These also happen to be exercises concerning the federal government for which the national percentages of success are low, 26.7%, 34.8% and 47.2%. They also share four exercises on which they are most varied from each other and from the national percentages. The national percentages are relatively high in these cases ranging from 67% to 82.5%. Two of the questions are about the federal government, and two are about local government functions. On two of these four, 81.7% and 87.4% of the high metro respondents indicated that the federal government carries mail and local governments carry garbage. Only 57.5% and 61.7% of the low metro students know these relationships. Main big city respondents are not much better; 65.8% know that garbage collection is most commonly a local government function. It is on that question that main big city respondents are most distinguished from other 13-year-olds excepting their neighbors in the inner city. Main big city and low metro responses to this question, which refers specifically to the likelihood of increasing garbage collections, may be more valid comments on city sanitary departments than on the knowledge levels of the respondents.

At age 17, all STOC groups except high and low metro, are clustered close to the national percentages of success. Perhaps a dozen delta p-values for the middle five groups are large enough to be significant. In contrast, all but five delta p-values for the low metro groups are significantly below the national percentages and all but two values for the high metro groups are significantly above the national percentages. For the most part, the two groups respond differently to the questions in this color scheme. As groups, they not only differ from each other but also from all other groups in the sample.

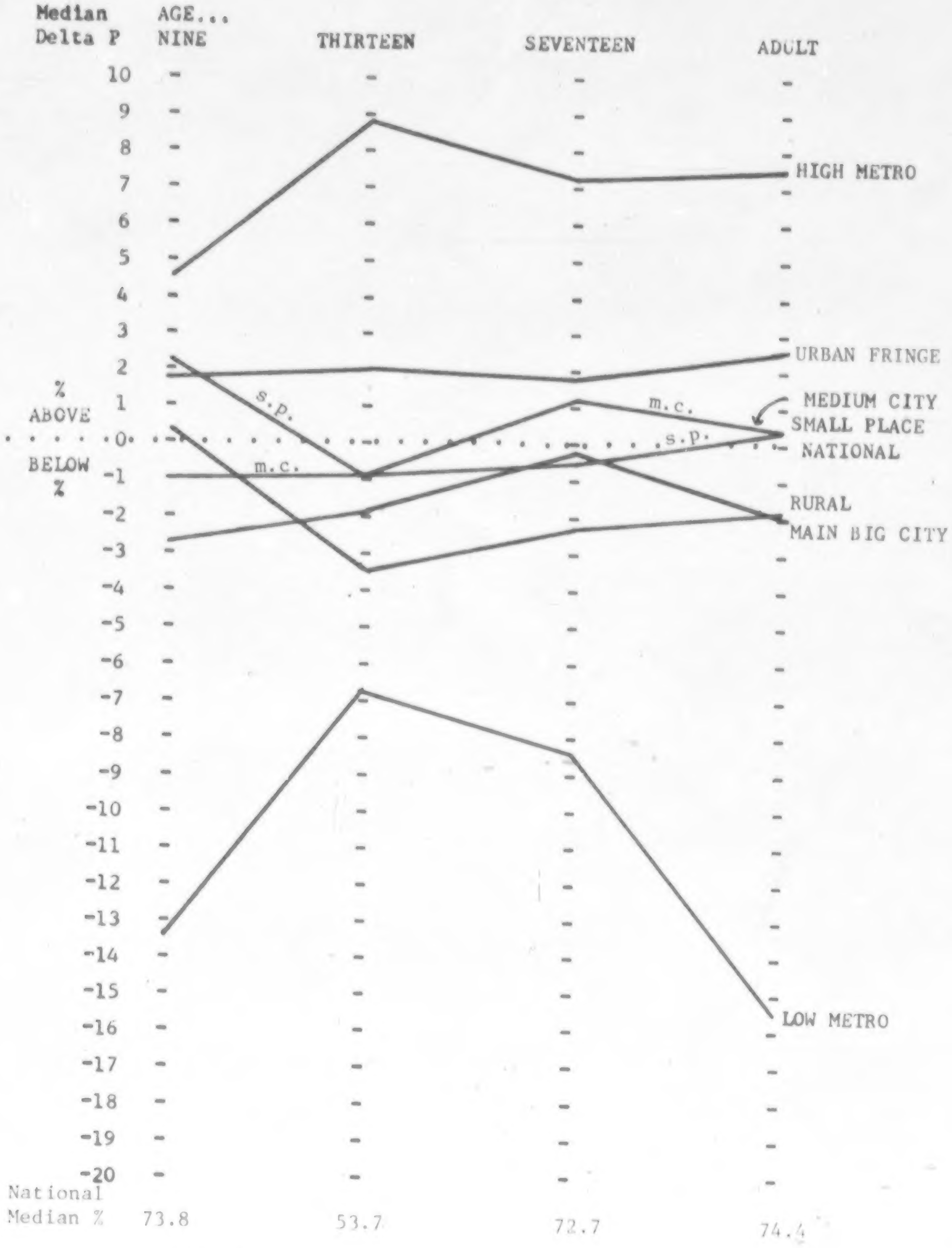


EXHIBIT 23. Knowledge of Institutional Structures - STOC - Social Studies 1971-72

On two questions, however, these two groups are alike. In fact, with the exception of the main big city group on one of these questions, the entire sample is homogeneous. On the ballot question, which asks if a voter can vote for both Kirk and Jones for a single Senate seat, 90% of the 17-year-olds said "no." On an unreleased question about school governance, about 65% of all 17-year-olds demonstrate that they know the locus of a certain decision. Only the main big city group is a significant 8.4% below the national percentage.

Three other released questions are illustrative of the difference between the high and low metro groups. Two of these have to do with the federal government. On a question about raising mail rates, 93.3% of the high metro group correctly identified this to be a federal government function; only 78% of the low metro group make this connection. When asked how a presidential candidate is nominated, 60% of the high metro group said it is by a national convention; only 29.7% of the low metro group know this procedure. At the local level, 95.3% of the high metro group know that local governments commonly collect garbage; but only 76.1% of the low metro 17-year-olds recognize this social task as a local government function.

As is observable in Exhibit 23, the relative positions of the groups do not change greatly from the 17-year-old sample to the adult sample. The high metro group is high, the low metro group is low, and the remaining five groups congregate closely about the median national percentage of success.

All high metro delta p-values are above the national percentages from 1.5% to 20.7%. All but two of these positive values are significant. All but two delta p-values of the urban fringe group are above the national percentages from 0.3% to 5.5%. The two values below the national percentages are -0.3% and -0.7%. About half of the positive values are significant. About half of the medium city delta p-values are above the national percentages from 0.4% to 4.7% and about half are below from -0.3% to -2.4%. None is significant. Over half of the delta p-values of the small place group are above the national percentages from 0.1% to 3.7%; the remainder are below from 0.1% to 4.6%. Three of those above and two of those below are significant. All but two of the extreme rural group's values are below the national percentages from -0.5% to -13.3%; one is at the national percentage and one is 1.5% above. Five of the negative values are significant. About one third of the delta p-values for the main big city group are above the national percentages from 0.4% to 6.9%; three are significant. The remaining two thirds are below from -0.5% to -13.7% with six significantly below. All of the low metro delta p-values are below the national percentages from -1.3% to -28.2%. All but two are significantly below.

As in the other age samples, a pattern emerges of a central core of five groups that behave homogeneously on these exercises. With the exception of a half dozen results each for the main big city, extreme rural and possibly the urban fringe groups, this core group deviates no more than ±5.0% from the national percentages on all exercises. The deviants are the high and low metro groups. With the exception of a half dozen results between them, these groups deviate no less than ±5.0% from the national percentages on all exercises.

The central core groups respond similarly on many exercises, but when the high and low metro groups are included, the adults respond alike on only one exercise. An unreleased question about the armed forces was answered correctly by 96.3% of all adults. No group deviated more than ±1.5% from the national percentage. The next nearest alike response is on the ballot exercise where respondents are asked if they can vote for both Kirk and Jones for a state's open Senate seat. No group deviated from the national percentage of 90.2% more than about 2.0% except for the low metro group. It is a significant -5.3% below the national percentage.

On another of the general election ballot questions the difference between the high and low metro performances is illustrated. Among all adults, 70% interpreted the ballot accurately as enabling them to vote for a councilman from each party. More than 80% of the high metro respondents did so; but only 46% of the low metro group made the correct interpretation.

On another exercise, about 60% of all adults indicated a knowledge of national conventions. Nearly 80% of the high metro group know of the procedure, but only about 30% of the low metro group said they know presidential candidates are nominated in conventions.

Nearly every exercise, it seems, is a potential example of the relative disadvantage of the low metro group in this assessment. The kinds of explanations one offers for the relatively poor performance of this group depend partially at least on his point of view and the s[illegible]f concepts that view entails. One point of view might depend heavily on a concept of intelligence; another might utilize broader concepts of cognitive abilities; still another would use concepts of culture and culture differences. The possible points of view are several and the concepts many.

One conclusion would necessarily have to be accounted for whatever explanation is hypothesized. Whatever is measured by these exercises is measured with different results in these polar groups. If some kind of ability attends a successful performance on these exercises — and that assumption seems a necessary one — then the high metro group is more in possession of it than the low metro group. What, exactly, that ability is is less easily described than we might wish. That it is associated with schoolwork seems a likely inference. That it is nurtured most effectively in an affluent environment seems also legitimately inferred. That it is better to have than not to have or that it is related to anything else worth having are nubbier assumptions. This assessment cannot help us with those.

*Rights and Duties.* Exhibit 24 displays the median delta p-values for the seven STOC categories for the Knowledge of Institutional Rights and Duties in the three age samples, 13, 17 and adult. There are too few exercises to allow the inclusion of the age 9 sample in these exhibits.

Exhibit 24 fairly adequately depicts the variance among these groups of 13-year-olds on the 10 exercises used to measure Knowledge of Institutional Rights and Duties. More specifically, however, we have tried to note on which exercises respondents seem to be quite similar across the sample and on which exercises they tend to be quite different. When we performed this analysis on the Structure questions, the high and low metro groups appeared dissimilar on most questions in all age samples. At the same time, the other five STOC groups appeared quite similar on many questions. In this color scheme, all 13-year-olds are alike relative to one exercise. When asked if there should be laws against vandalism, 93.1% of all 13-year-olds indicated that there should be. No group deviated more than 1.4% from this national percentage. None of the deviations are significant.

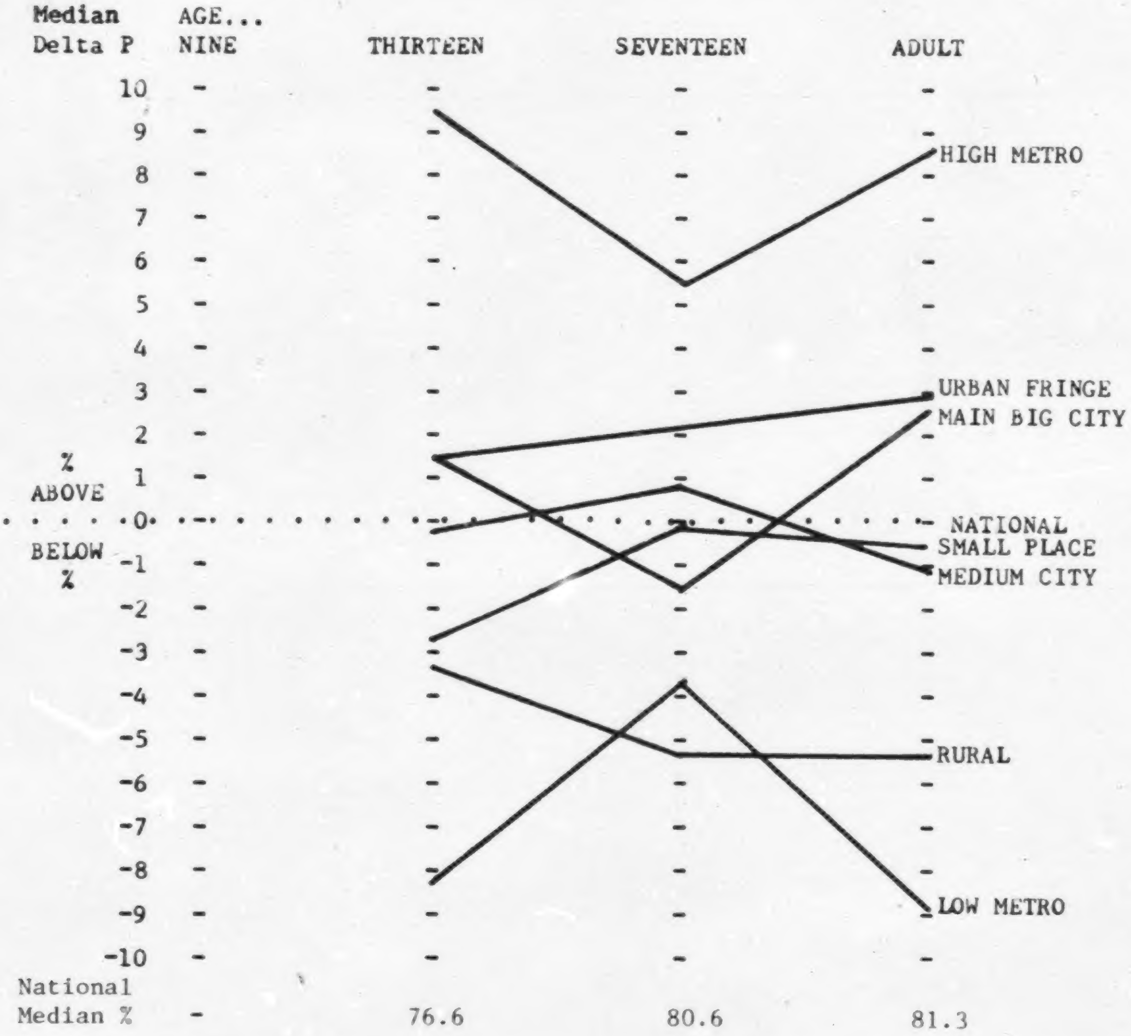


EXHIBIT 24. Knowledge of Institutional Rights and Duties - STOC - Social Studies 1971-72

On two other questions, one on due process and the other on free assembly, five of the STOC groups are homogeneous. But 13-year-olds in the high metro group are significantly above these others by 7.5% and 10.5% and respondents from the small place group are significantly below their colleagues by -3.5% and -4.3%.

On three exercises related to religious freedom, the 13-year-olds are not alike. For example, the 13-year-olds in the high metro and main big city groups are more likely to support for office a person who does not believe in God. Nearly 60% of all 13-year-olds would allow such a person to hold office; 75% of the high metro group and about 63% of the main big city youths are of this mind. At the same time, only 55% of the 13-year-olds from small places and 46% of the low metro youths support this belief. On the other two unreleased religious freedom questions, the variations are different, though the high metro group registers consistently higher percentages than any other group. The extreme rural, low metro and small place groups are significantly below the national percentages on one or the other of the exercises.

Four of the STOC groups, medium city, urban fringe, main big city and low metro, show no significant variance from the national percentage of 88.8% on the question of racial discrimination in employment. The other three groups, however, vary significantly on this question — small place residents by -3.2%, extreme rural by -4.1% and high metro by 7.4%.

Fewer than half of all 13-year-olds believe in or recognize the freedom of the press in politically sensitive areas. Both extreme rural and low metro groups are even below this level at 36% and 39.3%. The high metro group is significantly above the national percentage and all other groups at 61.5%. Freedom of the press gets short shrift among 13-year-olds, as do several other basic democratic rights and freedoms.

By 17 years of age an overall gain of 4% is noted relative to the 13-year-old median national percentage. To the extent that the two sets of exercises are comparable, the gain may be interpreted as an increase in the understanding of our institutional rights and duties or possibly an increase in commitment to them. But also, as can most easily be seen in Exhibit 24, there are some shifts in the relative position of some groups. The most notable of these, according to an analysis of exercises, is the exchange of extreme rural and low metro groups in the position of exhibiting the lowest median delta p score. The main big city and high metro groups also shift downward. However, because the main big city's median delta p-value stays clustered with those closest to national percentages, its change in status does not make so much difference on most exercises. Also, since the high metro group retains its highest position relative to a national percentage, its distribution is not much different.

Within the 17-year-old sample, the high and low metro and extreme rural groups are most deviant. We will emphasize these groups in the discussion and refer incidentally to the other groups, which tend to be fairly homogeneous.

In an earlier section we referred to a set of two questions as seeming to have a law and order orientation. It included the often-cited exercise on vandal laws and an unreleased question on crime reporting. In this age sample, extreme rural youths record their highest delta p-values on these exercises. The small place group almost equals the extreme rural youths' special concern for these matters. At the opposite end, the high metro group places these two questions in low priority. It achieves significant negative delta p-values on both. That in itself is notable, for the high metro group has very few delta p-values below the national percentages.

More characteristically, the high metro group has positive delta p-values ranging from 12% to nearly 16% on a set of five exercises that seem to measure preferences for open protest, freedom of the press and freedom of religion. We suggested earlier that these questions have a liberal orientation. The 17-year-olds as a total group are not overly fond (or apprised) of these ideas, but rural and small place youths find them especially distasteful (or unfamiliar). About 45% of the small place group and 35% of the rural group, for example, would permit a police station to be picketed. On this exercise the low metro achieves its highest delta p-value; 57.8% of that group would allow such a protest. Low metro youths do not tend to extend this right to include picketing of rock festivals, however; only 46.2% would allow that action. The high metro group does not differentiate between these objects; nearly two thirds would allow both a police station and a rock festival to be picketed.

The hardest question for 17-year-olds tested their knowledge of the Supreme Court's decision on religion in schools. Only half marked it as relating to the principle of church and state separation. The urban fringe group did best of all groups on that question; nearly 60% could answer it correctly. That amounts to a 20% advantage over the low metro group, which did the poorest of all groups on the question.

The easiest question for 17-year-olds relates to race as a factor in employment. It is also one of the few questions on which the 17-year-olds are much alike. Except for the main big city group, which produced a delta p-value of -4.3%, all groups are within 2% of the national percentage of 93.8%.

At the adult level, the most interesting comparisons are again among the high and low metro and extreme rural groups. The interest derives from their extreme variance on a number of exercises. Those variances are reflected in the median values reported in Exhibit 24. The four central groups, urban fringe, main big city, small place and medium city, vary from the national percentages on individual exercises no more than ±8.2%. In contrast, the extreme rural adults have a dispersion of 25.5% running from 3.8% to -21.7%; the low metro group goes from 6.4% to -20.8% for a dispersion of 27.2%; and the high metro group has a low delta p-value of 0.0% and a high value of 25.2%.

Nine out of 20 exercises in this color scheme have been released. As a means of illustrating the wide variances among these three groups, 6 of these 9 exercises are listed below in Table 10 along with their national percentages of success and the delta p-values for the high and low metro and extreme rural groups. An * beside a delta p-value indicates that the difference from the national percentage is large enough to be statistically significant.

The point of the display in Table 10 is to illustrate that the adults in these three STOC groups are indeed different in the way they perform on the Rights and Duties exercises. The groups also appear to vary from each other on the remaining 14 exercises in the color scheme. The extreme rural and low metro groups come closest together on the question concerning the right to assemble. Only 4 percentage points separate the results. On three exercises they are separated by about 10 percentage points, and on one they are 28.1 percentage points apart. The three groups are most nearly alike in their responses on the question about racial discrimination. On that question, the high and low metro groups are alike, i.e., their deviations from the national percentage are not significant.

### TABLE 10. Comparison of High and Low Metro and Extreme Rural Adults on Selected Rights and Duties Exercises — Social Studies 1971-72, Delta P

| Short Text of Exercise | National Percentage | Extreme Rural Delta P | High Metro Delta P | Low Metro Delta P |
|---|---|---|---|---|
| Allow picketing of police station | 52.2 | -21.7* | 16.5* | 6.4* |
| Supreme Court decision on religion in schools | 52.3 | -8.8* | 17.4* | -18.3* |
| Allow atheist to hold public office | 55.6 | -14.4* | 25.2* | -3.8 |
| Allow picketing of rock festival | 63.0 | -10.6* | 15.1* | -20.8* |
| Assemble in park for protest | 77.5 | -13.8* | 10.3* | -9.8* |
| Race a factor in employment | 92.3 | -6.3* | 0.2 | 2.3 |

The high metro and extreme rural groups are most different in their responses to the question on picketing police. Over two thirds of the high metro adults would allow that; only one third of the extreme rural adults would. The high and low metro groups are most different in their responses to picketing rock festivals. Where 78.1% of the high metro adults would support such an activity, only 42.2% of the low metro adults would.

An implication of these differences is that more than a single factor affected the responses of the three groups. Picketing as an activity seems relatively equally acceptable to high metro adults in very different contexts. It seems fairly unacceptable to extreme rural adults in either circumstance, but more so in one than in the other. It seems more acceptable to low metro adults when directed against police than when directed against rock festivals. A social comment with several facets seems buried in those results.

## Summary

We attempted as an overall strategy to divide a selected set of 48 exercises from the national assessment of social studies of 1971-72 into two parts. One part is comprised of items that we judged to deal mainly with various structural elements of institutions. The other part is comprised of items that we judged to deal primarily with rights and duties within these institutions. Though we persist in using the broader term of "institutions," most of the exercises, as it turns out, relate most directly to the institution of government. Some of the same exercises were used as the basis of the NAEP publication, *Political Knowledge and Attitudes 1971-1972*.[2]

In retrospect, the division of exercises into Structure and Rights and Duties color schemes appears to have had only a visual warrant since the two sets of questions produced very similar statistical discriminations among the various categories of persons assessed. That makes one believe that similar factors operated in the two parts; the two color schemes, in effect, appear to have assessed the same thing, whatever that is. We could continue to insist obdurately — which we do — that the subject matter division is valid, visually if not otherwise, and that the samples of persons assessed are in roughly equal possession of the two contents.

The parallel results along with either of these explanations can be thought of as partial justification for another decision made about the content of the two color schemes. Both sets of questions were approached as if they measured knowledge. That decision will not bear up under a visual inspection of exercises, however, because some questions, especially in the Rights and Duties color scheme, contain the value term "should" as the interrogative. Our *a priori* justification for largely ignoring this elementary distinction in these questions is argued elsewhere. Mainly, that justification suggests that the involved questions refer to fully institutionalized rights in this society and do not, in a real sense, pose issues. These questions, the argument continues, may just as validly be assessing these persons' knowledge of these rights as their preferences for them. While *Political Knowledge and Attitudes* treats some of these questions as outright value questions, as do we guardedly in several of our discussions, the parallel results of the two-scheme analysis used in this report seem to add some measure of justification to ignoring the distinction for this assessment.

An assumption was invoked on occasion in this analysis that results across age samples imply growth or change. The use of line graphs to depict the median performances of the categories of persons at different age levels appears to endorse this assumption. The truth of the matter is that we do not know for certain whether the performance of any age level in this assessment can be projected either forward or backward on the basis of the performance of any other age level, even though they may share several questions. Therefore, all statements that suggest or state outright that the 17-year-olds learned something of this content between the ages of 13 and 17 or that the adults have not learned much of this information since their school days must be read more cautiously than they were written. We found it useful to hypothesize these projections, but more and different information than is available from this assessment would be needed to say such things reliably.

Even with all the analyses, categories, color schemes and age samples, it is still most difficult to account for any of the results in the assessment. The reasons for this interpretive impasse are so well known by thoughtful school people that it seems gratuitous to mention them. Interpretive impasse notwithstanding, occasionally we did cast what could pass for an explanatory hypothesis; albeit, at some later date we may need to seek forgiveness for some of them. What we have here is a mass of descriptive data, confounded in all the ways that such data are always confounded, in spite of NAEP's meticulosity, and in which exists not one explanation.

A few content generalizations appear to emerge with some consistency in the themes reported in this analysis. One that comes to mind is the recurrent emphasis on law and order and the protection of property. Whatever the reasons for the high acceptance of these values, their emphasis in this assessment by many groups suggests something other than a general breakdown in the concern for security in this society. Another emphasis that seems clear is the high agreement by most groups that race should not be a factor in employment. That would seem to represent a major change in the social preferences in this country over the past decade or so.

We, of course, can state several facts with some firmness about the categories of persons assessed. We can say that persons of all ages — i.e., ages 9, 13, 17 and adult — whose parents went to college, persons of all ages who live in relative affluence, persons who live in the Northeast part of our country, males at age 17 and as adults, and whites at all ages are the most able of all categories of persons sampled to respond according to the criteria of this assessment.

We can also say, given the content, definitions and criteria of this assessment, that persons whose parents never went beyond the eighth grade, persons who live in the inner city, persons who live in the Southeast part of this country, females and blacks do not score as well on these exercises as do persons in most other categories. The data tell these facts, but they don't answer the important questions.

The important questions fall in two categories. In the first of these, the questions stab angrily into the inner mechanisms and justifications of NAEP. Other researchers in this social studies project have raised some of these important questions. Others have been raised in different forums. This report has perhaps raised fewer than it should. One stands out as signally important at the moment. Is it possible to construct an assessment that examines validly, reliably and fairly a plural population? There seems obviously a sense in which pluralism should dictate the standards that are applied in such a mass accounting of our schools and ourselves. Our differences and diversity surface in this social studies assessment as they do in all such assessments. But we seem easily tempted to judge the reasonable responses of others who enjoy different advantages than our own as less reasonable than our own.

In another category, the important questions seem mostly to test our social understandings and motivations. Why do these groups of persons behave differently on this assessment? The assessment's yield of descriptive data won't suffice to account for the differences. Do we want — more importantly, should we want — such groups to perform more evenly? If we do, how can schools help to make it happen?

The two categories of questions turn rapidly into each other. While we assume that some of the diversity exhibited in the assessment is a function of the assessment and some a function of schooling, beyond these, much of it is undoubtedly a product of this society. If for some reasons we desire a population that performs more evenly in such assessments, alterations that go beyond the unbiasing of a test and that go beyond the equalizing of schooling are likely to be necessary. Given the apparent value conflicts in this society and the different frames of reference that are displayed by this population, such a desire would seem empty, even if it were justifiable.

## *NOTES*

1. See Appendix D, Reference 4, pp. 9, 40-41.
2. See Appendix D, Reference 12.

# CHAPTER 9

# A RESPONSE FROM NATIONAL ASSESSMENT

J. Stanley Ahmann

The 36 social studies educators who participated in the NCSS review of the first assessments in citizenship and social studies are to be commended for their thoughtful analysis and suggestions for improvement. The report provides valuable input for decisions to be made in the years ahead. In commissioning the study, NAEP's purpose was twofold: to bring the assessments to the attention of the social studies education community via a thorough appraisal by key personnel from its own ranks, and to lay the foundation for further cooperation between our two organizations. We feel that these goals have been achieved although there are still hurdles to be surmounted before all will be satisfied with the end results.

The National Assessment of Educational Progress (NAEP) has a primary purpose to monitor change over time through the collection of education statistics at the national level. This longitudinal aspect means that the 1969 citizenship and the 1971 social studies assessments provide only baseline data, although some hypotheses may be drawn as several of the authors have suggested. Because the objectives on which these assessments were based resulted from a broad consensus, they reflect what the majority, at the time, considered desirable goals and behavior in these two areas for the majority of the population. This fact should not disturb those who do not hold these same goals nor should it be construed as saying anything about the merits of alternate philosophies. Measurement of all phases of a pluralistic society is a financially impossible feat, and National Assessment does not attempt to do this. Because of the dynamic nature of the NAEP model, however, there will be ample opportunity for changes in the nature of these objectives whenever the majority so determines.

In the dozen years since the concept of a national assessment was first debated, ideas, comments, criticisms and compliments have come to the project. To date almost a thousand scholars, educators and laymen, plus almost half a million youth and young adults have participated in the development of objectives, the review of exercises and the actual surveys in the 10 learning areas. As the second cycle of assessments begins, a number of major changes are in progress.

During the past two years the federal goverment has cut funds for HEW projects, including NAEP. With the current economic situation the NAEP budget is not likely to return to its peak period in fiscal 1973 for sevaral years. As a result, certain modifications of the original model have been made. For example:

1. The 10 learning areas have been consolidated into 5 — science, mathematics, reading, humanities (art, music and literature), and citizenship/social studies. Writing and career and occupational development (COD) will be assessed as special probes in the sense that the assessments will be limited to subsets of the field — i.e., writing will be limited to writing mechanics and the special probe in COD will be administered only to 17-year-olds.
2. Beginning in fiscal 1975 the adult sample has been dropped, and starting in fiscal 1976 (when the citizenship/social studies reassessment will occur) the out-of-school 17-year-olds will also be eliminated from the sample.
3. The total number of exercise packages must be reduced although both individual and group exercises will continue to be administered.
4. The number of selective reports for each learning area will be limited to three.
5. Staff expansion has been curtailed and the practice of subcontracting in all areas other than data collection and scoring has been eliminated.

How will the reassessment of citizenship/social studies differ from the original assessments? The citizenship and social studies objectives were reviewed in 1969 and 1972, respectively, preparatory to the development of new exercises for the second-cycle assessments, which were to have taken place in 1974 for citizenship and 1977 for social studies. Readers will note that a number of the concerns expressed in the NCSS study were anticipated by those engaged in this review. The following excerpts from the revised objectives booklets summarize the nature of the changes. A complete list of both the original and revised objectives and subobjectives for each learning area can be found in Appendixes A and B.

## Citizenship

The main changes in the citizenship objectives prepared for the 1974-75 assessment may be summarized as follows:

1. For some subobjectives, reviewers identified important omissions at the 9- and 13-year age levels.

Subobjectives and appropriate illustrative behaviors have been added for these age levels in several instances.

2. The substantive nature of civic problems was left implicit in the original objectives for the most part, except that international problems were spelled out in detail. The substantive description of major civic problems which a citizen needs to understand in order to act effectively has been expanded. In addition to extending problems in social conflict to include local and national problems, two other major problem areas affected by civic policy, namely, economic needs (poverty, employment, etc.) and environmental problems (pollution, etc.) have been specified.
3. In response to criticism that the objective on knowledge of government was too limited to textbook ideals, illustrative behaviors were added concerning knowledge of informal influences on government, sources of actual power and bureaucracy. Effectiveness of citizen participation rather than effort alone was also emphasized.
4. Objectives dealing with personal development and voluntary personal relations were in many instances concluded to be too remotely related to citizenship to retain. These aspects of citizenship, while not eliminated, received less emphasis in the revised objectives.
5. The behaviors listed under a number of the original objectives seemed to emphasize middle-class values which might not be accepted as goals by other social strata of the nation. In several instances such behaviors were deleted or changed to examples with more universal appeal. For example, "controlling emotions in the face of criticism" was changed to "express emotions in nondestructive ways."

The objectives are intended to be a working guide for the difficult task of assessment, not a description of how an ideal citizen should spend his day. No one person could be expected to exhibit all of the specific behaviors included. But since the assessment is intended to describe the achievements of a wide population of citizens, not individual persons, this presents no problem. For example, it might be reported that "10 percent of 17-year-old boys in the country have served as leaders of a group engaged in civic activities." Such a result would not be intended to disparage in any way those 17-year-olds who felt they were not qualified to act in such a role or to imply that "the more time a person spends at civic pursuits, the better."

National Assessment objectives are not permanent standards of achievement; rather, they are intended to reflect the evolution of goals in education in response to the changing needs of the nation.[1]

## Social Studies

During the winter of 1972-73, the social studies objectives delineated in this booklet were formulated by social studies and social science specialists from various universities and secondary and elementary schools, and by lay people in various occupations from differents parts of the country.

One of the challenges in the process of developing objectives for the next assessment of social studies has been the necessity for subject-matter specialists and lay people to develop a framework of major objectives that would allow for the development of more specific subobjectives and related age-specific illustrations. It was decided that an appropriate framework for the measurement of achievement in the social studies would include the acquisition of a knowledge base for understanding human beings and their relationships with their environments, and understanding of values as they relate to individuals and groups, the use of intellectual and human-relation skills, the development of a positive self-concept and a sense of and commitment to rational social participation.

> The knowledge to be acquired in social studies draws on the content of the major social science disciplines but is organized in terms of interdisciplinary concepts and ideas to allow a broader and more integrated approach to the study of social phenomena. The skills to be acquired are those of intellectual inquiry and human relations that enable the learner to ask questions about social problems and to participate rationally and responsibly in society. An understanding of the values of individuals and groups are included in recognition of the need to explore and clarify value orientations that underlie our institutions and those of other societies. It is also appropriate to recognize that a person's self-perception is closely related to the individual's role in society. Since social studies deals with the study of the individual's role in society, the development of a positive self-concept becomes an important objective in this area. The importance of the participation of young people in the solution of societal problems prompted the inclusion of a final objective that addresses itself to the commitment to the right of self-determination for all human beings and a willingness to take rational action in support of means for securing and preserving human rights.[2]

When the decision was made to combine citizenship and social studies in one assessment under the revised time schedule, the citizenship redevelopment process had been virtually completed. The social studies redevelopment had only progressed as far as the stage of objectives review. Therefore, the 1975-76 assessment will be predominantly a citizenship one plus a reassessment of the original unreleased social studies exercises. New citizenship exercises will reflect the recommendations of the Hunkins committee for greater emphasis on upper cognitive skills and more exercises at the affective level. Furthermore, future social studies exercises are expected to take into account the results of a survey of social studies educators concerning the amount of weight each subobjective and objective should carry. While a subobjective may still have only a limited number of exercises, by choice, the overall objective will have ample coverage.

Specifically, the 1975-76 citizenship/social studies assessment will consist of 49 new citizenship exercises, 55 original citizenship and 22 original social studies exercises. When overlapped administrations across the three in-school age levels are counted, the result will be approximately 360 exercises. One fourth of the new exercises and two thirds of the old ones will be released after the assessment. The remainder will be the basis of the comparison data for the third assessment.

Since the chief function of NAEP is to measure change it is imperative that exercises used for comparsion purposes be identical to their original form. Although a few unreleased exercises have been dropped from the second cycle, the bulk of them will reappear. While this action may discourage those who would prefer to see some changes in the wording of certain exercises, it should be noted that no exercise is used that has not passed an extensive screening process. The NAEP exercise-review process has been increased fourfold from that used for the first assessment. It is anticipated that the exercises that survive this screening will meet with greater approval.

What does the future hold for NAEP? Citizenship and social studies will remain a single assessment, but attention will need to be given to a blending of the two. Furthermore, among the related ongoing research efforts by staff is an examination of other background data that might be collected about the individuals participating in the survey, which could aid those who must make decisions about the findings. One such study has already resulted in the monograph *Associations Between Educational Outcomes and Background Variables: A Review of Selected Literature*, and staff are engaged in developing a plan that will incorporate some new features — but not for the 1975-76 assessment.

Documentation has been completed regarding the individual responses for all released social studies exercises from the first assessment and may be obtained by independent researchers from the ED STAT II division of the National Center for Education Statistics in Washington, D.C. Other computer tapes already on file there include NAEP reading the literature responses. Citizenship files from the first assessment are not available, but citizenship-/social studies files from the second assessment will be added as soon as possible after the completion of the data analysis.

Hopefully, conditions will one day allow the resumption of data collection for adults and out-of-school 17-year-olds. Meanwhile, we shall continue to call upon professional associations to help identify strengths and weaknesses in National Assessment. This NCSS report clearly demonstrates the helpful information that can result from a thorough evaluation of assessment data and the objectives and exercises behind them. The entire education community should benefit from the careful work of these members of the National Council for the Social Studies.

## *NOTES*

1. See Appendix D, Reference 2, pp. 4-5.
2. See Appendix D, Reference 15, pp. v-vi.

# APPENDIX A

## CITIZENSHIP OBJECTIVES

### 1969-70 Assessment

**I. Show Concern for the Welfare and Dignity of Others**
   - A. Treat all individuals with respect.
   - B. Consider the consequences for others of their own actions.
   - C. Guard safety and health of others.
   - D. Help other individuals voluntarily.
   - E. Are loyal to country, to friends, and to other groups whose values they share.
   - F. Understand and oppose unequal opportunity in the areas of education, housing, employment, and recreation.
   - G. Seek to improve the welfare of groups of people less fortunate than they.

**II. Support Rights and Freedoms of All Individuals**
   - A. Understand the value of constitutional rights and freedoms.
   - B. Recognize instances of the proper exercise or denial of constitutional rights and liberties, including due process of law.
   - C. Defend rights and liberties of all kinds of people uniformly.

**III. Help Maintain Law and Order**
   - A. Understand the need for law and order.
   - B. Are conscious of right and wrong behavior.
   - C. Comply with public law and school rules.
   - D. Help authorities in specific cases.
   - E. Protest unjust rules openly.
   - F. Inform themselves about the law.

**IV. Know the Main Structure and Functions of Our Governments**
   - A. Recognize the purposes of government.
   - B. Recognize the main functions and relations of governmental bodies.
   - C. Recognize the importance of political opposition and diverse interest groups.
   - D. Recognize that democracy depends on the alertness and involvement of its citizens, and know how citizens can affect government.
   - E. Recognize the structure and operation of political parties.
   - F. Know structure of school and student government.

**V. Seek Community Improvement Through Active, Democratic Participation**
   - A. Believe that each person's civic behavior is important, and convey this belief to others.
   - B. Recognize important civic problems and favor trying to solve them.
   - C. Actively work for community improvement.
   - D. Participate in local, state and national governmental processes.
   - E. Apply democratic procedures on a practical level when working in a group.
   - F. Display fairness and good sportsmanship toward others.

**VI. Understand Problems of International Relations**
   - A. Are aware of the problems of international conflict and dangers to national security.
   - B. Seek world peace and freedom for all peoples.

**VII. Support Rationality in Communication, Thought and Action on Social Problems**
   - A. Try to inform themselves on socially important matters and to understand alternative viewpoints.
   - B. Evaluate communications critically and form their own opinions independently.
   - C. Weigh alternatives and consequences carefully, then make decisions and carry them out without undue delay.
   - D. See relations among social problems and have good ideas for solutions.
   - E. Support free communication and communicate honestly with others.
   - F. Understand the role of education in developing good citizens.

**VIII. Take Responsibility for Own Personal Development and Obligations**
   - A. Further their own self-improvement and education.
   - B. Plan ahead for major life changes.
   - C. Are conscientious, dependable, self-disciplined, and value excellence and initiative.
   - D. Economically support self and dependents.

**IX. Help and Respect Their Own Families (Ages 9, 13, 17)**

- A. Respect the reasonable authority of their parents, or guardians, and help with home duties and problems.
- B. Help younger brothers and sisters to develop into good citizens.
- C. Discuss social matters with their families and respect the views of all family members.

**X. Nurture the Development of Their Children As Future Citizens (Adults)**

- A. Provide for the basic needs and health of their children.
- B. Encourage cooperative, ethical relations to authority and to other individuals.
- C. Develop in their children a broadening awareness, independence, and rationality.

## SOCIAL STUDIES OBJECTIVES

## 1971-72 ASSESSMENT

**I. Have Curiosity About Human Affairs**

- A. Identify and define problems and issues.
- B. Formulate generalizations and hypotheses capable of being tested.
- C. Obtain information from a variety of sources.
- D. Distinguish facts from opinion, relevant from irrelevant information, and reliable from unreliable sources.
- E. Detect logical errors, unstated assumptions, and unwarranted assertions; question unsupported generalizations; are aware of the complex nature of social causation and understand that sequence or relationship does not necessarily imply causation.
- F. Use data and evaluative criteria to make decisions.

**II. Use Analytic-Scientific Procedures Effectively**

- A. Raise questions and seek answers.
- B. Are open to new information and ideas.
- C. Try to understand why other people think and act as they do.

**III. Are Sensitive to Creative-Intuitive Methods of Explaining the Human Condition**

- A. Read history, philosophy, and fiction.
- B. Obtain insight into human affairs from history and philosophy, and from fiction and other forms of art.
- C. Recognize the role of creative-intuitive methods in scientific inquiry.
- D. Distinguish personalized explanations of human affairs from scientific-objective explanations.

**IV. Have Knowledge Relevant to the Major Ideas and Concerns of Social Scientists**

- A. Understand some of the distinctive modes of inquiry (questions and approaches) of social scientists.
- B. Understand some of the major relationships involving culture, the group, and the self.
- C. Understand some of the major characteristics of economic systems, expecially the American economic system.
- D. Understand some of the major characteristics of the geographic (spatial) distributions of man and his activities, and of man's interaction with the physical environment.
- E. Understand some of the major historical developments.
- F. Understand some of the characteristics of the major systems of government, particularly the political system of the United States.

**V. Have a Reasoned Commitment to the Values That Sustain a Free Society**

- A. Believe in the fundamental worth of the individual and can justify their belief.
- B. Believe in the freedoms of the First Amendment and can justify their belief.
- C. Believe in the rule of law and can justify their belief.
- D. Believe in open opportunity for advancement and can justify their belief.
- E. Are willing to act for the general interest.
- F. Are willing to participate in decision making relevant to their lives.

# APPENDIX B

## REVISED CITIZENSHIP OBJECTIVES

**I. Show Concern for the Well-Being and Dignity of Others**

A. Treat others with respect.

B. Consider the consequences for others of their own actions.

C. Guard safety and health of others.

D. Offer help to others in need.

E. Support equal opportunity in education, housing, employment, and recreation.

F. Are loyal to country, to friends, and to other groups whose values they share.

G. Are ethical and dependable in work, school, and social situations.

**II. Support Just Law and the Rights of All Individuals**

A. Understand the need for law.

B. Recognize specific constitutional rights and liberties.

C. Defend rights and liberties of all kinds of people.

D. Encourage ethical and lawful behavior in others.

E. Comply with public laws.

F. Oppose unjust rules, laws, and authority by lawful means.

**III. Know the Main Structure and Functions of Their Governments**

A. Recognize basic governmental purposes.

B. Understand the organization of federal and state governments.

C. Know the political structure of their community.

D. Recognize the relationships of different levels of government.

E. Recognize the importance of political opposition and interest groups.

F. Recognize that democracy depends on the alertness and involvement of its citizens, and know how citizens can affect government.

G. Know structure of school and student government.

**IV. Participate in Democratic Civic Improvement**

A. Believe that each person's civic bvehavior is important, and convey this belief to others.

B. Favor organized civic action where it is needed.

C. Actively work for civic improvement.

D. Participate in local, state, and national governmental processes.

E. Apply democratic procedures effectively in small groups.

**V. Understand Important World, National, and Local Civic Problems**

A. Understand social conflict among individuals, groups, and nations and the difficulties in achieving peace and social harmony.

B. Recognize how different civic policies may affect people's efforts to meet their economic needs.

C. Recognize major environmental problems and are aware of alternative civic solutions.

D. See relations among civic problems and particular events.

E. Generate good ideas about causes and solutions for civic problems.

**VI. Approach Civic Decisions Rationally**

A. Seek relevant information and alternative viewpoints on civically important decisions.

B. Evaluate civic communications and actions carefully as a basis for forming and changing their own views.

C. Plan and organize civic tasks effectively.

D. Support open, honest communication and universal education.

**VII. Help and Respect Their Own Families**

A. Cooperate in home responsibilities and help provide for other family members.

B. Instill civic values and skills in other family members.

# REVISED SOCIAL STUDIES OBJECTIVES

## 9-YEAR-OLDS

**I. Develops a knowledge base for understanding the relationships between human beings and their social and physical environment.**

- A. Acquires knowledge about social organization.
- B. Acquires knowledge about the relationships between human beings and their social environments and understands some of the consequences of these relationships.
- C. Acquires knowledge about the relationships between human beings and their physical environments and understands some of the consequences of these relationships.
- D. Acquires knowledge about decision-making processes.
- E. Acquires knowledge about conflict and the impact it has on individual and group relationships.

**II. Develops an understanding of the origins and interrelationships of beliefs, values and behavior patterns.**

- A. Expresses awareness of some of the beliefs and values expressed by people and recognizes that the conditions, times and places in which people live influence their beliefs, values and behaviors.
- B. Understands ways beliefs and values are transmitted in various cultures.
- C. Understands some of the influences of differing beliefs and values on relationships between people.
- D. Examines own beliefs and values and the interrelationships between and among beliefs, values and behavior.

**III. Develops the competencies to acquire, organize and evaluate information for purposes of solving problems and clarifying issues.**

- A. Identifies problems or issues appropriate for study.
- B. Prepares a plan to guide study of a problem or issue.
- C. Identifies, locates and uses sources of information and evaluates the reliability and relevance of these sources.
- D. Organizes, analyzes, interprets and synthesizes information obtained from various sources.
- E. Uses summarized information to draw conclusions, offer solutions to problems or clarify issues.
- F. Validates outcome of study.

**IV. Develops the human relation skills necessary to communicate and work with others.**

- A. Attends to expressions of others.
- B. Encourages others to express views and opinions.
- C. Listens carefully to others.
- D. Clarifies and elaborates on own ideas.
- E. Asks for clarification and elaboration of the ideas of others.
- F. Expresses awareness of different discussion roles (e.g., initiator, facilitator, blocker) and recognizes some of the effects of these roles on individual and group action.
- G. Interacts in various capacities (e.g., leader, advisor, supporter).
- H. Expresses willingness to interact with a variety of people.
- I. Provides emotional and intellectual support for others in group efforts.
- J. Shares in responsibilities that arise from group efforts.

**V. Develops a positive self-concept, builds self-esteem and moves toward self-actualization.**

- A. Expresses awareness of the characteristics that give one identity.
- B. Expresses awareness of one's goals (aspirations), the goals of the groups with which one identifies and the fit between these goals.
- C. Expresses awareness of the relative strengths of oneself and the groups with which one identifies and recognizes the societal barriers to full development that may exist.
- D. Assesses the extent to which one has control over the setting and achievement of personal goals in light of what one knows about oneself, the groups with which one identifies and the societal barriers to full development.
- E. Suggests ways to maximizing one's effectiveness.

**VI. Develops and demonstrates a commitment to the right of self-determination for all human beings and a willingness to take rational action in support of means for securing and preserving human rights.**

- A. Displays an awareness of a quality of human life and an interest in ways in which the quality can be improved.
- B. Explains and supports rights and freedoms important to human development.
- C. Participates in family, school and community life on the basis of rational decisions involving one's own values and the conflict among these values.

## 13-YEAR-OLDS

**I. Develops a knowledge base for understanding the relationships between human beings and their social and physical environment.**

- A. Acquires knowledge about social organization.
- B. Acquires knowledge about the relationships between human beings and their social environments and understands some of the consequences of these relationships.

C. Acquires knowledge about the relationships between human beings and their physical environments and understands some of the consequences of these relationships.
D. Acquires knowledge about decision-making processes.
E. Acquires knowledge about conflict and the impact it has on individual and group relationships.

**II. Develops an understanding of the origins and interrelationships of beliefs, values and behavior patterns.**
A. Expresses awareness of some of the beliefs and values expressed by people and recognizes that the conditions, times and places in which people live influence their beliefs, values and behaviors.
B. Understands ways beliefs and values are transmitted in various cultures.
C. Understands some of the influences of differing beliefs and values on relationships between people.
D. Examines own beliefs and values and the interrelationships between and among beliefs, values and behavior.

**III. Develops the competencies to acquire, organize the evaluate information for purpose of solving problems and clarifying issues.**
A. Identifies problems or issues appropriate for investigation.
B. Plans how to investigate a problem or issue.
C. Identifies, locates and uses sources of information and evaluates the reliability and relevance of these sources.
D. Organizes, analyzes, interprets and synthesizes information obtained from various sources.
E. Uses summarized information to draw conclusions, offer solutions to problems, clarify issues or make predictions.
F. Validates outcomes of investigation.
G. Appraises judgments and values that are involved in the choice of a course of action.

**IV. Develops the human relation skills necessary to communicate and work with others.**
A. Attends to expressions of others.
B. Encourages others to express views and opinions.
C. Listens carefully to others.
D. Clarifies and elaborates on own ideas.
E. Asks for clarification and elaboration of the ideas of others.
F. Expresses awareness of different discussion roles (e.g., initiator, facilitator, blocker) and recognizes some of the effects of these roles on individual and group action.
G. Interacts in various capacities (e.g., leader, advisor, supporter).
H. Expresses willingness to interact with a variety of people.
I. Provides emotional and intellectual support for others in group efforts.
J. Shares in responsibilities that arise from group efforts.

**V. Develops a positive self-concept, builds self-esteem and moves toward self-actualization.**
A. Expresses awareness of the characteristics that give one identity.
B. Expresses awareness of one's goals (aspirations), the goals of the groups with which one identifies and the fit between these goals.
C. Expresses awareness of the relative strengths of oneself and the groups with which one identifies and recognizes the societal barriers to full development that may exist.
D. Assesses the extent to which one has control over the setting and achievement of personal goals in light of what one knows about oneself, the groups with which one identifies and the societal barriers to full development.
E. Suggests ways of maximizing one's effectiveness.

**VI. Develops and demonstrates a commitment to the right of self-determination for all human beings and a willingness to take rational action in support of means for securing and preserving human rights.**
A. Displays an awareness of a quality of human life and an interest in ways in which the quality can be improved.
B. Explains and supports rights and freedoms important to human development.
C. Participates in family, school and community life on the basis of rational decisions involving one's own values and the conflict among these values.

## 17-YEAR-OLDS AND ADULTS

**I. Develops a knowledge base for understanding the relationship between human beings and their social and physical environment.**
A. Acquires knowledge about social organization.
B. Acquires knowledge about the relationships between human beings and their social environments and understands some of the consequences of these relationships.
C. Acquires knowledge about the relationships between human beings and their physical environments and understands some of the consequences of these relationships.
D. Acquires knowledge about decision-making processes.
E. Acquires knowlede about conflict and the impact it has on individual and group relationships.

**II. Develops an understanding of the origins and interrelationships of beliefs, values and behavior patterns.**
A. Expresses awareness of some of the beliefs and values expressed by people and recognizes that the conditions, times and places in which people live influence their beliefs, values and behaviors.
B. Understands ways beliefs and values are transmitted in various cultures.

C. Understands some of the influences of differing beliefs and values on relationships between people.

D. Examines own beliefs and values and the interrelationships between and among beliefs, values and behavior.

**III. Develops the competencies to acquire, organize and evaluate information for purposes of solving problems and clarifying issues.**

A. Identifies problems or issues appropriate for investigation.

B. Designs a plan to investigate a problem or issue.

C. Identifies, locates and uses sources of information and evaluates the reliability and relevance of these sources.

D. Organizes, analyzes, interprets and synthesizes information obtained from various sources.

E. Uses summarized information to draw conclusions, offer solutions to problems, clarify issues, make predictions or serve as a guide to continued investigation.

F. Validates outcomes of investigation.

G. Appraises judgments and values that are involved in the choice of a course of action.

**IV. Develops the human relation skills necessary to communicate and work with others.**

A. Attends to expressions of others.

B. Encourages others to express views and opinions.

C. Listens carefully to others.

D. Clarifies and elaborates on own ideas.

E. Asks for clarification and elaboration of the ideas of others.

F. Expresses awareness of different discussion roles (e.g., initiator, facilitator, blocker) and recognizes some of the efforts of these roles on idividual and group action.

G. Interacts in various capacities (e.g., leader, advisor, supporter).

H. Expresses willingness to interact with a variety of people.

I. Provides emotional and intellectual support for others in group efforts.

J. Shares in responsibilities that arise from group efforts.

**V. Develops a positive self-concept, builds self-esteem and moves toward self-actualization.**

A. Expresses awareness of the characteristics that give one identity.

B. Expresses awareness of one's goals (aspirations), the goals of the groups with which one identifies and the fit between these goals.

C. Expresses awareness of the relative strengths of oneself and the groups with which one identifies and recognizes the societal barriers to full development.

D. Assesses the extent to which one has control over the setting and achievement of personal goals in light of what one knows about oneself, the groups with which one identifies and the societal barriers to full development.

E. Suggests ways of maximizing one's effectiveness.

**VI. Develops and demonstrates a commitment to the right of self-determination for all human beings and a willingness to take rational action in support of means for securing and preserving human rights.**

A. Displays an awareness of a quality of human life and an interest in ways in which the quality can be imporved.

B. Explains and supports rights and freedoms important to human development.

C. Participates in family, school and community life on the basis of rational decisions involving one's own values and the conflict among these values.

# APPENDIX C

# EVALUATION OF PERFORMANCE ON SELECTED SOCIAL STUDIES EXERCISES

This appendix covers only a portion of the released exercises and is not intended to be a representative sample. A complete report on all released exercises may be found in *The First Social Studies Assessment: An Overview*, listed in Appendix D. The panel's task (see Chapter 7) was not to predict what performance was to be but to judge what performance (realistically) ought to be. The panel's judgments are no more, but no less, than those of nine competent social studies educators. Readers will do well to ask themselves with what performance levels they are satisfied.

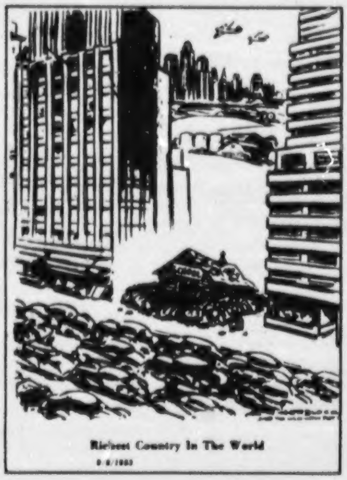
This appendix is divided into two parts. The first provides examples of exercises covering knowledge, skills and attitudes in general. The second section lists all the released exercises used by the Cox committee (see Chapter 8). The Cox committee also reviewed 22 unreleased exercises.

Key to exercise numbering system: 1st letter — R: released; 2nd letter — S: skills, A: attitudes, K: knowledge; 3rd letter — O: obtaining information, I: interpreting information, R: rights of the 1st amendment, B: worth of the individual, E: economics, G: geography, H: history, P: political science; 4th and 5th numbers are the numbers of the exercise within that category; 6th-9th letters would be subquestions within a multiple-part exercise.

| Sample Exercises from the First Social Studies Assessment | Nat'l Level of Acceptable Performance: Actual (Percent) |  |  |  | Realistically Satisfactory Performance Level: Panel (Percent) |  |  |  |
|---|---|---|---|---|---|---|---|---|
|  | Age 9 | Age 13 | Age 17 | Adult | Age 9 | Age 13 | Age 17 | Adult |
| RS120CD |  |  |  |  |  |  |  |  |
| The American Declaration of Independence states:<br>"We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty, and the pursuit of Happiness — That to secure these rights, Governments are instituted among Men, deriving their just powers from the consent of the governed . . ."<br>Do the following sentences restate ideas expressed in this quotation from the Declaration of Independence? |  |  |  |  |  |  |  |  |
| C. Some rights can never be rightfully taken away.<br>Yes<br>No<br>I don't know | — | 75 | 80 | 78 | — | 61-80 | >80 | >80 |
| D. Governments get their right to govern people from the people.<br>Yes<br>No<br>I don't know | — | 80 | 90 | 94 | — | 61-80 | >80 | >80 |
| RKE16 |  |  |  |  |  |  |  |  |
| A major American manufacturing corportation seeks to establish a branch plant in a country that has rich natural resources but very little industry. The leaders of the nation turn down the American corporation's request.<br>What reasons can you give for the decision made by the leaders of the foreign nation? | — | — | 52 | 52 | — | — | 61-80 | 61-80 |
| RKH13ABCD |  |  |  |  |  |  |  |  |
| American Indians, Black Americans, Oriental Americans and Spanish-speaking Americans have contributed a great deal to the history and culture of our nation. For each of the groups I read tell me the names of as many famous or nationally known men and women as you can. The person named may be either living or dead. Briefly describe each person's contribution or field of work. Consider people in ANY field of work — the Arts, Business, Civil Rights, Education, Entertainment, Politics, Science or Sports. |  |  |  |  |  |  |  |  |
| A. American Indians | 6 | 18 | 35 | 41 | 61-80 | 61-80 | 61-80 | 61-80 |
| B. Black Americans | 8 | 34 | 64 | 73 | 61-80 | >80 | >80 | >80 |
| C. Oriental Americans | 0 | 0 | 1 | 4 | 41-60 | 61-80 | 61-80 | 61-80 |
| D. Spanish-speaking Americans | 1 | 3 | 7 | 20 | 61-80 | 61-80 | 61-80 | 61-80 |
| RKH18 |  |  |  |  |  |  |  |  |
| Which one of the following is the MAJOR goal of the United Nations?<br>To fight disease<br>To maintain peace<br>To spread democracy<br>To fight the Communists<br>I don't know | 47 | 77 | 92 | 89 | 41-60 | 61-80 | 61-80 | 61-80 |

| Sample Exercises from the First Social Studies Assessment | Nat'l Level of Acceptable Performance: Actual (Percent) | | | | Realistically Satisfactory Performance Level: Panel (Percent) | | | |
|---|---|---|---|---|---|---|---|---|
| | Age 9 | Age 13 | Age 17 | Adult | Age 9 | Age 13 | Age 17 | Adult |
| RKE04<br>Billy went to a department store with his mother. As they went into the store, Billy saw a sign in the store window. The sign looked like this:<br>BUY NOW<br>PAY LATER | | | | | | | | |
| Which one of the following tells what the sign means?<br>The store is having a sale.<br>Some things in the store are free.<br>You can pay at another time for what you buy today.<br>You can trade something you have for something you want.<br>I don't know. | 89 | — | — | — | 61-80 | — | — | — |
| RKE09BC<br>Which of the following things happen when a country becomes highly industrialized? | | | | | | | | |
| B. There is greater emphasis on individual craftsmanship.<br>Yes<br>No<br>I don't know | — | 41 | 66 | 53 | — | 41-60 | 61-80 | 61-80 |
| C. There is a movement of people from rural to urban parts of the country.<br>Yes<br>No<br>I don't know | — | 67 | 78 | 83 | — | 61-80 | 61-80 | 61-80 |
| RKE07<br>Economists divide purchases into two groups: producer goods and consumer goods. A farmer buys seed, fertilizer, a tractor and a new coat for his wife. | | | | | | | | |
| Which one of the things that the farmer bought is in the group of consumer goods?<br>Seed<br>Tractor<br>Fertilizer<br>New coat for his wife<br>I don't know | — | 57 | 81 | 79 | — | 61-80 | 61-80 | >80 |
| RKG05<br>Which one of the following states borders on the Atlantic Ocean?<br>California<br>Nebraska<br>New York<br>Ohio<br>I don't know | 37 | — | — | — | 61-80 | — | — | — |

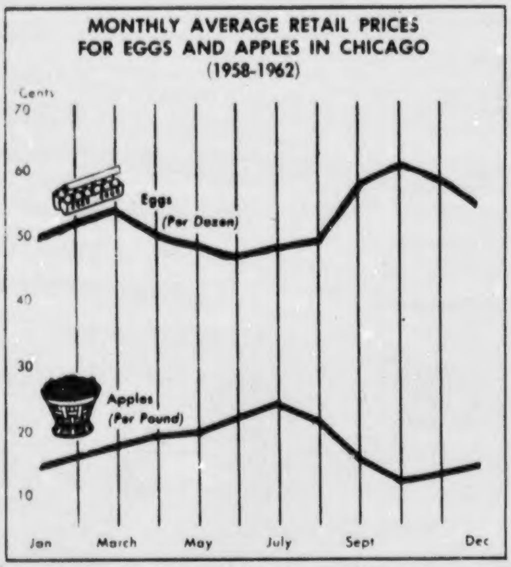
| Sample Exercises from the First Social Studies Assessment | Nat'l Level of Acceptable Performance: Actual (Percent) | | | | Realistically Satisfactory Performance Level: Panel (Percent) | | | |
|---|---|---|---|---|---|---|---|---|
| | Age 9 | Age 13 | Age 17 | Adult | Age 9 | Age 13 | Age 17 | Adult |
| RKE12<br>The term "monopoly" describes the situation in which the market price of goods and services is established by which one of the following?<br>Many sellers<br>A single buyer<br>Many buyers and sellers<br>A single seller or a small group of sellers<br>I don't know | — | — | 21 | 30 | — | — | 61-80 | 61-80 |
| RSI31<br>Look at the cartoon. What idea is the artist trying to put across in this cartoon? | — | 83 | 88 | 91 | — | 61-80 | >80 | >80 |


Richest Country In The World

Cartoon by Herbert Block, "Richest Country in the World," from *Herblock's Here and Now* (New York: Simon & Schuster, 1955), p. 273. Reprinted by permission of Herbert Block.

| Sample Exercises from the First Social Studies Assessment | Nat'l Level of Acceptable Performance: Actual (Percent) | | | | Realistically Satisfactory Performance Level: Panel (Percent) | | | |
|---|---|---|---|---|---|---|---|---|
| | Age 9 | Age 13 | Age 17 | Adult | Age 9 | Age 13 | Age 17 | Adult |
| RS007ABCDE<br>A. What do you think are three important problems facing large cities in the United States?<br>B. Which one of the problems you named would you MOST want to ask questions about?<br>C. What two questions would you ask about this problem to find out more about it?<br>D. Name two sources that would help you learn more about the problem you most want to ask questions about.<br>E. Name as many additional sources as you can that would help you learn more about the problem you MOST want to ask questions about. | — | 38 | 59 | 66 | — | 41-60 | 61-80 | >80 |

| Sample Exercises from the First Social Studies Assessment | Nat'l Level of Acceptable Performance: Actual (Percent) Age 9 | Age 13 | Age 17 | Adult | Realistically Satisfactory Performance Level: Panel (Percent) Age 9 | Age 13 | Age 17 | Adult |
|---|---|---|---|---|---|---|---|---|
| RS014CD<br>Suppose someone wanted to know what had happened in our country or in the world within the last 24 hours. Would each of the following be a good way for that person to find out? | | | | | | | | |
| C. Would reading the daily newspaper be a good way to find out what had happened in the last 24 hours? | 87 | 98 | — | — | >80 | >80 | — | — |
| D. Would looking in an encyclopedia be a good way to find out what had happened in the last 24 hours? | 89 | 98 | — | — | >80 | >80 | — | — |
| RSI23<br>Look at the graph, then answer the question below it. | | | | | | | | |



Reprinted by permission of the Bureau of Labor Statistics

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| The average retail price of apples was highest in which month?<br>January<br>July<br>October<br>December<br>I don't know | — | 89 | 96 | 91 | — | 61-80 | >80 | >80 |

| Sample Exercises from the First Social Studies Assessment | Nat'l Level of Acceptable Performance: Actual (Percent) |  |  |  | Realistically Satisfactory Performance Level: Panel (Percent) |  |  |  |
|---|---|---|---|---|---|---|---|---|
|  | Age 9 | Age 13 | Age 17 | Adult | Age 9 | Age 13 | Age 17 | Adult |
| RSI33<br>Please listen carefully to this recording of "Carefully Taught" from *South Pacific*. While listening, try to identify the main idea or message.<br>You've got to be taught<br>To hate and fear<br>You've got to be taught<br>From year to year<br>It's got to be drummed<br>In your dear little ear<br>You've got to be carefully taught<br>You've got to be taught<br>To be afraid<br>Of people whose eyes are oddly made<br>And people whose skin is a different shade<br>You've got to be carefully taught.<br>You've got to be taught before it's too late<br>Before you are 6 or 7 or 8<br>To hate all the people<br>Your relatives hate.<br>You've got to be carefully taught<br>You've got to be carefully taught.<br>What is the song about? | — | 22 | 39 | 39 | — | 61-80 | >80 | >80 |
| RS011<br>A boy looked in his history book, but he could not find out where Abraham Lincoln was born. Which one of the following should he do?<br>Look in an atlas<br>Look in an encyclopedia<br>Look in a geography<br>Ask a friend to help him<br>I don't know | 68 | — | — | — | 61-80 | — | — | — |
| RS117C<br>Some things CAN be proved to be true or false, some things CANNOT BE proved to be true or false.<br>Read each of the statements below and decide whether it can or cannot be proved. If you think it CAN be proved, fill in the oval beside "Can be proved." If you think it CANNOT be proved, fill in the oval beside "Cannot be proved." If you do not know the answer, fill in the oval beside "I don't know." An example is ~~done for you~~.<br>*Example*<br>The earth is almost round.<br>Can be proved<br>Cannot be proved<br>I don't know<br>C. People from Mexico are nicer than people from Canada.<br>Can be proved<br>Cannot be proved<br>I don't know | 53 | 71 | 79 | 72 | 61-80 | 61-80 | >80 | >80 |

| Sample Exercises from the First Social Studies Assessment | Nat'l Level of Acceptable Performance: Actual (Percent) | | | | Realistically Satisfactory Performance Level: Panel (Percent) | | | |
|---|---|---|---|---|---|---|---|---|
| | Age 9 | Age 13 | Age 17 | Adult | Age 9 | Age 13 | Age 17 | Adult |
| RAB11<br>Suppose a friend from India comes to your house for dinner. Your mother is making hamburgers for dinner. While you are playing, your friend tells you he does not eat meat. His religion will not let him eat meat. What should you do? | 87 | — | — | — | 61-80 | — | — | — |
| RAB04<br>Do you think the people who live in a neighborhood should be allowed to decide who can and cannot live in their neighborhood? | — | — | 80 | 63 | — | — | >80 | >80 |
| RKP01<br>Below are listed four of the many jobs that are done in a city. Which one of the jobs is done by the health department?<br>Selling food<br>Directing traffic<br>Putting out fires<br>Inspecting restaurants<br>I don't know | 36 | — | — | — | 61-80 | — | — | — |
| RKP02<br>In a court, which one of the following has the job of making sure that the trial is fair and run according to the rules?<br>The judge<br>The lawyer<br>The jury<br>The person on trial<br>I don't know | 74 | — | — | — | 61-80 | — | — | — |
| RKP16<br>In the United States which one of the following men is elected to office?<br>A United States senator<br>The United States secretary of state<br>A United States Supreme Court justice<br>The United States ambassador to Great Britain<br>I don't know | — | 74 | 89 | 90 | — | 61-80 | 61-80 | >80 |
| RKP17<br>The presidential candidate for each major political party is formally nominated by which one of the following?<br>The Senate<br>A national primary<br>A national convention<br>The House of Representatives<br>I don't know | — | 17 | 49 | 60 | — | 61-80 | 61-80 | 61-80 |
| RKP13<br>Which one of the following has the power to declare an act of Congress unconstitutional?<br>The Congress<br>The president<br>The United States Supreme Court<br>The United States Department of Justice<br>I don't know | — | 35 | 71 | 62 | — | 61-80 | 80 | 61-80 |

| Sample Exercises from the First Social Studies Assessment | Nat'l Level of Acceptable Performance: Actual (Percent) | | | | Realistically Satisfactory Performance Level: Panel (Percent) | | | |
|---|---|---|---|---|---|---|---|---|
| | Age 9 | Age 13 | Age 17 | Adult | Age 9 | Age 13 | Age 17 | Adult |
| RKP03<br>The head of government in the United States is the president. Which one of the following is usually the head of government in a town?<br>The mayor<br>The governor<br>The chief of police<br>The school principal<br>I don't know. | 58 | — | — | — | 41-60 | — | — | — |

RKP18

The ballot below was used in a general election. Look at the ballot to answer the questions on this and the following two pages.

| | LEGISLATIVE | | COUNTY | | |
|---|---|---|---|---|---|
| OFFICES | SENATOR IN CONGRESS (vote for one) | REPRESENTATIVE IN CONGRESS (vote for one) | COUNCILMAN (vote for two) | | TAX ASSESSOR (vote for one) |
| DEMOCRATIC | Alan F KIRK | John G SMITH | Martha G DAVIS | Peter V MOSS | |
| REPUBLICAN | James M JONES | Mary O'CONNOR | John RICHARDS | Michael M MERWIN | Joseph L LASKI |

| | Age 9 | Age 13 | Age 17 | Adult | Age 9 | Age 13 | Age 17 | Adult |
|---|---|---|---|---|---|---|---|---|
| A. If you wanted to vote for Kirk for senator, could you also vote for O'Connor for member of the House of Representatives?<br>Yes<br>No<br>I don't know | — | — | 41 | 44 | — | — | >80 | >80 |
| B. Could you vote for both Davis and Moss for councilman? | | | | | | | | |
| C. Could you vote for both Davis and Merwin for councilman? | | | | | | | | |
| D. If you were registered as a member of the Democratic Party, could you vote for Laski for tax assessor? | | | | | | | | |
| E. Could you vote for both Kirk and Jones for senator? | | | | | | | | |
| RKP08B<br>Which one of the following would MOST likely pass an act to raise the rates for sending letters through the mail?<br>Federal government<br>State government<br>Local government<br>I don't know | — | 72 | 90 | 95 | — | 61-80 | 61-80 | >80 |
| RKP08C<br>Which one of the following would MOST likely pass an act to lower taxes on goods coming into the country?<br>Federal government<br>State government<br>Local government<br>I don't know | — | 73 | 89 | 92 | — | 61-80 | >80 | >80 |

| Sample Exercises from the First Social Studies Assessment | Nat'l Level of Acceptable Performance: Actual (Percent) | | | | Realistically Satisfactory Performance Level: Panel (Percent) | | | |
|---|---|---|---|---|---|---|---|---|
| | Age 9 | Age 13 | Age 17 | Adult | Age 9 | Age 13 | Age 17 | Adult |
| RKP08D<br>Which one of the following would MOST likely pass an act to increase garbage collection services?<br>Federal government<br>State government<br>Local government<br>I don't know | — | 77 | 92 | 92 | -- | 61-80 | >80 | >80 |
| RKP11<br>The Supreme Court ruled that it is unconstitutional to require prayer and formal religious instruction in public schools. Which one of the following was the basis for its decision?<br>The requirements violated the right to freedom of speech.<br>There was strong pressure put on the Supreme Court by certain religious minorities.<br>Religious exercises violated the principles of the separation of church and state.<br>Every moment of the valuable school time was needed to prepare students to earn a living.<br>I don't know. | — | — | 39 | 52 | — | — | >80 | 61-80 |
| RAB17<br>Should a congressman pay attention to the opinions and concerns of people whose views are different from those of the majority?<br>Yes<br>No<br>I don't know<br>Please explain any answer you selected. | — | — | 79 | 78 | — | — | 61-80 | 61-80 |
| RAR05<br>Indicate whether you agree or disagree with the following statement: Anyone who criticizes the use of United States troops in military action abroad should be prohibited from expressing his views publicly.<br>Agree<br>Disagree<br>Please explain your position. | — | — | 81 | 81 | — | — | >80 | >80 |
| RAR11<br>In the picture, there are many people gathered together in a public park. They are demanding changes which you do not agree with. Should these people be allowed to gather and make their demands in a public place?<br>Yes<br>No<br>Undecided<br>No response | — | 54 | 80 | 72 | — | 61-80 | >80 | >80 |

| Sample Exercises from the First Social Studies Assessment | Nat'l Level of Acceptable Performance: Actual (Percent) | | | | Realistically Satisfactory Performance Level: Panel (Percent) | | | |
|---|---|---|---|---|---|---|---|---|
| | Age 9 | Age 13 | Age 17 | Adult | Age 9 | Age 13 | Age 17 | Adult |
| RAR08<br>Should a newspaper or magazine be allowed to publish something that criticizes an elected government official?<br>Yes<br>No<br>Undecided | — | 41 | 66 | 71 | — | >80 | >80 | >80 |
| RAR07<br>Should a person who does not believe in God be allowed to hold a public office?<br>Yes<br>No<br>Undecided | — | 59 | 63 | 56 | — | 61-80 | >80 | >80 |
| RAR12A<br>Do you think people should be allowed to picket the holding of a rock festival as a protest against it?<br>Yes<br>No<br>Undecided | — | — | 42 | 56 | — | — | >80 | >80 |
| RAR12B<br>Do you think people should be allowed to picket a police station to protest reported police brutality?<br>Yes<br>No<br>Undecided | — | — | 40 | 44 | — | — | >80 | >80 |
| RAB20<br>Do you think there should be laws against acts of vandalism such as destroying a statue?<br>Yes<br>No<br>Undecided | — | 55 | 51 | — | — | >80 | >80 | — |
| RAB16<br>Should race be a factor in hiring someone for a job?<br>Yes<br>No<br>I don't know | — | 74 | 90 | 90 | — | >80 | >80 | >80 |

# APPENDIX D

# BIBLIOGRAPHY OF NATIONAL ASSESSMENT PUBLICATIONS CITED

1. *Citizenship Objectives*, 1969-70 National Assessment of Citizenship. Ann Arbor: Committee on Assessing the Progress of Education, 1969.
2. *Citizenship Objectives*, 1974-75 National Assessment of Citizenship. Denver: National Assessment of Educational Progress, 1972.
3. *Citizenship: Group Results for Parental Education, Color, Size and Type of Community, Citizenship, Report 9*, 1969-70 Assessment. Washington, D.C.: Government Printing Office, 1972.
4. *Citizenship: Group Reusults for Sex, Region, Size of Community, Citizenship, Report 6*, 1969-70 Assessment. Washington, D.C.: Government Printing Office, 1971.
5. *Citizenship: National Results, Citizenship, Report 2*, 1969-70 Assessment. Washington, D.C.: Government Printing Office, 1970.
6. *Citizenship, National Results — Partial: Observations and Commentary of a Panel of Reviewers, Report 2*, 1969-70 Assessment. Washington, D.C.: Government Printing Office, 1970.
7. Finley, Carmen J., and Berdie, Frances S., *The National Assessment Approach to Exercise Development*. Ann Arbor: National Assessment of Educational Progress, 1970.
8. *First Social Studies Assessment: An Overview, Social Studies, Report 03-SS-00*, 1971-72 Assessment. Washington, D.C.: Government Printing Office, 1974.
9. *General Information Yearbook, Reading and Literature, Report 02-GIY*, 1970-71 Assessment. Washington, D.C.: Government Printing Office, 1972.
10. *NAEP Newsletter*. Denver: January-February 1974.
11. *NAEP Newsletter*. Denver: May-June 1974.
12. *Political Knowledge and Attitudes, Social Studies, Report 03-SS-01*, 1971-72 Assessment. Washington, D.C.: Government Printing Office, 1973.
13. *Science: National Results, Science, Report 1*, 1969-70 Assessment. Washington, D.C.: Government Printing Office, 1970.
14. *Social Studies Objectives*, 1971-72 National Assessment of Social Studies. Ann Arbor: National Assessment of Educational Progress, 1970.
15. *Social Studies Objectives*, 1975-76 National Assessment of Social Studies. Denver: National Assessment of Educational Progress, 1974.
16. Womer, Frank B., *What Is National Assessment?* Ann Arbor: National Assessment of Educational Progress, 1970.

U.S. GOVERNMENT PRINTING OFFICE: 1974 — 679-970 / 227 REGION NO. 8